



**UNIVERSIDADE FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS  
DEPARTAMENTO DE LINGÜÍSTICA**

**ANNA CRISTINA BARATIERI**

**A PRODUÇÃO DE OCLUSIVAS SURDAS EM LÍNGUA INGLESA: A  
INTELIGIBILIDADE DE APRENDIZES BRASILEIROS E O  
RECONHECIMENTO DE FALA**

**CURITIBA  
2017**

ANNA CRISTINA BARATIERI

**A PRODUÇÃO DE OCLUSIVAS SURDAS EM LÍNGUA INGLESA: A  
INTELIGIBILIDADE DE APRENDIZES BRASILEIROS E O  
RECONHECIMENTO DE FALA**

Dissertação apresentada como requisito parcial  
à obtenção do grau de Mestre em Letras, no  
Programa de Pós-Graduação em Letras, da  
Universidade Federal do Paraná.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Denise Cristina Kluge

CURITIBA

2017



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
Setor CIÊNCIAS HUMANAS  
Programa de Pós-Graduação LETRAS

### TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em LETRAS da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **ANNA CRISTINA BARATIERI** intitulada: **A PRODUÇÃO DE OCLUSIVAS SURDAS EM LÍNGUA INGLESA: A INTELIGIBILIDADE DE APRENDIZES BRASILEIROS E O RECONHECIMENTO DE FALA.**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 04 de Setembro de 2017.

DENISE CRISTINA KLUGE

Presidente da Banca Examinadora (UFPR)

GUSTAVO NISHIDA

Avaliador Externo (UTFPR)

RONALD BARRY MARTINEZ

Avaliador Interno (UFPR)

“Que quantidade imensa de coisas um homem deve fazer  
para que digamos que ele pensa.”

Ludwig Wittgenstein

## RESUMO

É cada vez mais corriqueiro o uso de aplicativos de reconhecimento de fala para realização de tarefas do dia-a-dia. Apesar de haver uma preocupação das empresas que mantêm esses aplicativos em coletar dados de fala compatíveis com seus usuários, os desenvolvedores dos aplicativos de reconhecimento de fala parecem não levar em conta os falantes não nativos dos idiomas disponíveis (OTHERO; MENUZZI, 2005), mesmo sabendo-se que mais da metade dos falantes de inglês são estrangeiros (CRYSTAL, 1992; ERLING, 2005; STATISTA, 2016), por exemplo. Entende-se, portanto, que sejam fundamentais estudos que reportem com que eficácia os recursos informáticos podem ser utilizados por falantes não nativos de Inglês, num contexto de Língua Franca. A presente investigação, pioneira na academia brasileira, então, debruça-se sobre a inteligibilidade do falante estrangeiro de Inglês para um aplicativo de reconhecimento de fala, disponível como ferramenta de transcrição automática no YouTube. Mais especificamente, a inteligibilidade de /p/, /t/ e /k/ em posição de ataque silábico produzidas por brasileiros aprendizes de Inglês. Para tal, participaram um grupo de estudantes de Curitiba (Paraná, Brasil), matriculados no curso de Inglês, respondendo a uma tarefa de familiaridade vocabular e a uma de produção. Além disso, um grupo de voluntários americanos e brasileiros atestou a inteligibilidade das amostras de fala coletadas. As análises estatísticas e descritivas permitiram perceber que há uma possível ordem de inteligibilidade entre os segmentos oclusivos surdos, em que [k] é mais inteligível que [p], que é mais inteligível que [t] tanto para a ferramenta de transcrição quanto para o grupo de ouvintes. Os dados também evidenciaram que as palavras-alvo produzidas pelos usuários brasileiros em que aparecem os segmentos /p, t, k/ em posição de ataque silábico são menos inteligíveis que o segmento em si. Ainda, pode-se afirmar que os segmentos /p/, /t/ e /k/ em posição de ataque silábico são mais inteligíveis em palavras-alvo familiares aos usuários brasileiros. Somado a isso, evidenciou-se que a frequência de uso das palavras-alvo possa ser mais impactante à inteligibilidade que a familiaridade dessas o são ao falante.

**Palavras-chave:** Fonética. Consoantes. Plosivas. Inteligibilidade. Reconhecimento de Fala.

## ABSTRACT

The use of voice recognition devices in daily activities is most frequent each day. Even though companies that maintain these apps show concern in gathering speech data compatible to its users, app developers seem not to acknowledge nonnative users of the available languages (OTHERO; MENUZZI, 2005), even knowing that more than half of English speakers are foreign (CRYSTAL, 1992; ERLING, 2005; STATISTA, 2016), for instance. It is understood as prior, then, studies that report to which efficiency computer resources can be used by nonnative English speakers, in a Lingua Franca context. The present investigation explores the intelligibility of the nonnative English speaker to a voice recognition device, available as an automatic transcription tool in YouTube. More specifically, the intelligibility of /p/, /t/ e /k/ in onset in the speech of Brazilian learners of English. Therefore, a group of students from Curitiba (Paraná, Brazil), enrolled in an English course, took part of the research answering to a familiarity task and a production one. Besides, a group of Brazilian and American volunteers warranted the intelligibility of the speech samples. Statistical and descriptive analysis allowed to notice a possible order in intelligibility among the English voiceless stops, in which [k] is more intelligible than [p], which is more intelligible than [t] for both the YouTube transcription tool and for the group of listeners. Data also showed that the target words produced by the Brazilian users are less intelligible than the target segments (/p/, /t/ e /k/) in onset. Still, it is also possible to say that /p/, /t/ e /k/ are more intelligible in target words that are familiar to the Brazilian users. Added to this, it was revealed that the usage frequency of the target words can be more relevant to intelligibility than the familiarity of them to the speaker.

**Palavras-chave:** Phonetics. Consonants. Stop segment. Intelligibility. *Recognition Intelligence.*

## LISTA DE FIGURAS

FIGURA 01 - Articulações bilabial (A), alveolar (B) e velar (C) (ROACH, 2009 apud SOUZA, 2012, p.16).....	29
FIGURA 02 – Forma de onda e espectrograma da produção da consoante [k] entre vogais do PB (BARBOSA, 2015, p.174).....	31
FIGURA 03 – Forma de onda e espectrograma da produção da consoante [g] entre vogais do PB (BARBOSA, 2015, p.175).....	32
FIGURA 04 – Distribuição dos valores de VOT para as aclusivas vozeadas e desvozeadas (KENT; READ, 2015, p.248).....	32
FIGURA 05 – Arquitetura de um sistema de reconhecimento de fala (ADAMI, 2010, p.5).....	40
FIGURA 06 – Relação entre minutagem e transcrição do <i>input</i> sonoro.....	65

## LISTA DE GRÁFICOS

GRÁFICO 01 – Familiaridade das palavras-alvo para o grupo de falantes.....	62
GRÁFICO 02 – Tempo (em segundos) demandado para a coleta das amostras de fala com o grupo de falantes.....	64
GRÁFICO 03 – Distribuição dos valores de inteligibilidade dos lexemas iniciados em [p] e [k] para a ferramenta de transcrição do YouTube.....	77
GRÁFICO 04 – Distribuição de inteligibilidade da consoante oclusiva alveolar por aprendiz brasileiro para a ferramenta de transcrição do YouTube.....	79



## LISTA DE QUADROS

QUADRO 01 – Pares mínimos das oclusivas surdas e sonoras em PB e IA....	18
QUADRO 02 – Características biográficas do grupo de voluntários para a pilotagem em 2015. ....	46
QUADRO 03 – Unidades lexicais presentes no material da instituição de ensino de línguas que atendem ao contexto /p, t, k + vogal anterior alta/.....	49
QUADRO 04 – Distribuição das palavras-alvo em frases. ....	50
QUADRO 05 – Palavras distratoras em função da consoante inicial.....	52
QUADRO 06 – Características biográficas do grupo de falantes.....	57
QUADRO 07 – Características biográficas do grupo de falantes cujas amostras de fala foram submetidas à transcrição automática do YouTube. ....	58
QUADRO 08 – Características biográficas do grupo de ouvintes.....	59
QUADRO 09 – Indicadores de progressão usados no experimento de coleta de amostras de fala.....	64
QUADRO 10 – Seleção de frases empregadas no teste de inteligibilidade. ....	67
QUADRO 11 – Familiaridade das palavras-alvo para o grupo de ouvintes. ....	68
QUADRO 12 – Distribuição quantitativa dos dados coletados. ....	70
QUADRO 13 – Média de inteligibilidade e distribuição dos segmentos oclusivos surdos produzidos pelo grupo de aprendizes brasileiros para a ferramenta de transcrição do YouTube.....	72
QUADRO 14 – Média de inteligibilidade da ferramenta de transcrição automática do YouTube e distribuição dos vocábulos iniciados por segmentos oclusivos surdos produzidos pelo grupo de aprendizes brasileiros. ....	72

QUADRO 15 – Médias de inteligibilidade dos segmentos oclusivos surdos e dos vocábulos iniciados por estes para a ferramenta de transcrição automática do YouTube.....	73
QUADRO 16 – Média de inteligibilidade do segmento oclusivo bilabial surdo /p/ para a ferramenta de reconhecimento do YouTube. ....	74
QUADRO 17 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo bilabial surdo /p/ para a ferramenta de reconhecimento do YouTube.....	75
QUADRO 18 – Transcrição do YouTube para as frases em que houve a inserção da fricativa [s] anteriormente ao segmento [p] na palavra-alvo <i>peach</i> .....	76
QUADRO 19 – Transcrição do YouTube para as frases em que houve o vozeamento da consoante /k/ na palavra-alvo <i>peaks</i> .....	77
QUADRO 20 – Média de inteligibilidade do segmento oclusivo alveolar surdo /t/ para a ferramenta de reconhecimento do YouTube.....	79
QUADRO 21 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo alveolar surdo /t/ para a ferramenta de reconhecimento do YouTube. ....	80
QUADRO 22 – Transcrição do YouTube para as frases em que houve o deslocamento da consoante /t/ para a posição de coda do vocábulo anterior à palavra-alvo <i>teens</i> . ....	81
QUADRO 23 – Transcrição do YouTube para as frases em que houve o deslocamento da consoante /t/ para a posição de coda do vocábulo anterior à palavra-alvo <i>team</i> . ....	81
QUADRO 24 – Média de inteligibilidade do segmento oclusivo velar surdo /k/ para a ferramenta de reconhecimento do YouTube. ....	82
QUADRO 25 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo velar surdo /k/ para a ferramenta de reconhecimento do YouTube.....	83

QUADRO 26 – Transcrição do YouTube para as frases em que houve a inserção da fricativa [s] anteriormente ao segmento [k] na palavra-alvo <i>keel/s</i> . .....	85
QUADRO 27 – Relação das frases transcritas pelos ouvintes referentes às amostras de fala dos falantes 01, 02, 03 e 10. ....	86
QUADRO 28 - Inteligibilidade do segmento-alvo referente cada palavra-alvo transcrita pelos quatro ouvintes. ....	86
QUADRO 29 - Inteligibilidade de cada palavra-alvo transcrita pelo grupo de quatro ouvintes. ....	88
QUADRO 30 – Transcrição da frase com a palavra-alvo <i>keel/s</i> pela ferramenta de transcrição do YouTube e pelo grupo de ouvintes. ....	90
QUADRO 31 – Relação descritiva dos fenômenos lingüísticos observados nas transcrições realizadas pela ferramenta do YouTube e pelo grupo de ouvintes.....	93

## LISTA DE SIGLAS

AI	- <i>Artificial Intelligence</i>
COCA	- <i>Corpus of Contemporary American English</i>
HMM	- <i>Hidden Markov Model</i>
IA	- Inglês Americano
ILE	- Inglês como Língua Estrangeira
ILF	- Inglês como Língua Franca
ISL	- Inglês como Segunda Língua
LFC	- <i>Lingua Franca Core</i>
L1	- primeira língua ou língua materna
L2	- segunda língua ou língua estrangeira
PB	- Português Brasileiro
SRI	- <i>Speech Recognition Intelligence</i>
VOT	- <i>Voice Onset Time</i>

## SUMÁRIO

PREFÁCIO .....	24
1 INTRODUÇÃO.....	15
1.1 JUSTIFICATIVA .....	17
1.2 OBJETIVOS .....	19
2 FUNDAMENTAÇÃO TEÓRICA .....	22
2.1 INTELIGIBILIDADE .....	23
2.2 AS OCLUSIVAS SURDAS .....	29
2.3 RECONHECIMENTO DE FALA .....	35
3 PROCEDIMENTOS METODOLÓGICOS .....	42
3.1 METODOLOGIA DO ESTUDO PILOTO .....	45
3.1.1 População alvo e seleção dos sujeitos do estudo piloto .....	45
3.1.2 <i>Corpus</i> do estudo piloto .....	48
3.1.3 Gravações e coleta de dados do estudo piloto .....	51
3.1.4 Limitações do piloto .....	53
3.2 METODOLOGIA DA COLETA DE DADOS DESTA PESQUISA .....	54
3.2.1 População alvo e seleção dos sujeitos .....	55
3.2.2 <i>Corpus</i> de coleta.....	60
3.2.3 Gravação e coleta de dados .....	60
4 APRESENTAÇÃO E ANÁLISE DOS DADOS .....	70
4.1 A INTELIGIBILIDADE PARA A FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE .....	70
4.1.1 A oclusiva surda bilabial .....	74
4.1.2 A oclusiva surda alveolar .....	78
4.1.3 A oclusiva surda velar.....	82
4.2 A INTELIGIBILIDADE PARA O GRUPO DE OUVINTES.....	85
5 CONSIDERAÇÕES FINAIS.....	92
REFERÊNCIAS.....	96
APÊNDICE A – TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO.....	101
APÊNDICE B – QUESTIONÁRIO BIOGRÁFICO PARA BRASILEIROS .....	105
APÊNDICE C – ROTEIRO DE ENTREVISTA.....	109
APÊNDICE D – QUESTIONÁRIO DE FAMILIARIDADE.....	111
APÊNDICE E – CARTÕES EMPREGADOS NO EXPERIMENTO DE FAMILIARIDADE .....	112

APÊNDICE F – LÂMINAS DE ORIENTAÇÃO PARA A COLETA DE AMOSTRAS SONORAS .....	117
APÊNDICE G – TUTORIAL PARA TRANSCRIÇÃO DE VÍDEOS NO YOUTUBE .....	119
APÊNDICE H – TRANSCRIÇÃO DAS AMOSTRAS DOS SUJEITOS 01, 02, 03 E 10 FEITA PELOS OUVINTES.....	120
APÊNDICE I – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 01 FEITA PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE .....	121
APÊNDICE J – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 02 FEITA PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE .....	124
APÊNDICE K – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 03 FEITA PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE .....	127
APÊNDICE L – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 10 FEITA PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE .....	130

## PREFÁCIO

Era muito provavelmente uma quarta-feira escaldante na fronteira com o Paraguai. Meu irmão e eu tentávamos manter nossa temperatura corporal estável enquanto tomávamos tererê.

Neste dia, com muito pouca disposição para qualquer tarefa que exigisse esforço físico, nos dispusemos a uma batalha de cordialidade com nossos dispositivos móveis, solicitando aos recursos de reconhecimento de voz de cada dispositivo que nos informasse o quanto bem se sentiam (*“Hey, Siri. How are YOU doing?”* – Joey.).

À época, havia uma limitação de configuração que não permitia o uso da língua portuguesa (nossa língua materna) com os recursos de reconhecimento de fala e, tendo o inglês como segunda língua, demos início à competição em língua estrangeira.

Não demorou para que constatássemos uma limitação no sistema de reconhecimento. Apesar de termos o mesmo modelo de dispositivo móvel configurado com a mesma versão de sistema operacional, a mesma estabilidade de conexão de Internet sem fio e estarmos localizados na mesma coordenada geográfica, Siri se negava a conversar com meu irmão.

Eu, na condição de irmã mais velha, absorvi as dores do caçula. Como ousava Siri não interagir com meu irmão, que tem, inclusive, boa aparência, alegando não o compreender? Eu entendia perfeitamente o que ele enunciava e, por isso, as respostas de Siri (*“Sorry, I couldn’t get that.”*) me pareciam evasivas.

Em resumo, poucas das estratégias que tentamos (diminuir velocidade de fala, lapidar pronúncia, etc.) se mostraram eficientes para o reconhecimento da fala do meu irmão. Em comparação, quando eu indagava Siri sobre seu bem-estar, ela respondia com disposição.

De alguma maneira, essa experiência que tinha como objetivo despretencioso o entretenimento, se expandiu para um pré-projeto de pós-graduação, submetido à avaliação do corpo docente da Universidade Federal do Paraná em 2014. O resto se descreve nas páginas que seguem.

## 1 INTRODUÇÃO

A linguística, há muito tempo, dedica-se ao estudo do processamento da linguagem humana. Desde o princípio, esforços foram empenhados para que fossem desenvolvidas teorias que explicassem como as línguas se organizam (PETTER, 2010; BARROS, 2010). Posteriormente, teóricos dedicaram seus estudos a propor comparações e contrastes entre as línguas naturais (SAUSSURE, 1916; CHOMSKY, 1957). Graças ao esforço desses pesquisadores, que somaram forças com outras áreas científicas, como a física e a medicina, foi possível entender mais sobre o universo linguístico e almejar outros feitos científicos que dependiam do conhecimento relacionado ao processamento e organização da linguagem. Dentre as teorias formuladas sobre a estruturação das línguas destaca-se a proposta por Chomsky (1957). Apresentada na década de 1950, a gramática gerativa pode ser encarada como um mecanismo de produção de frases de uma dada língua, e é chamada gerativa porque sugere que com um número finito de regras, pode-se gerar um número infinito de sequências sentenciais.

A gramática proposta por Chomsky (1957) é entendida tanto no nível mental quanto no pragmático. No nível mental, consiste na representação estrutural de como o léxico e as regras sintáticas de formação de sentenças se organizam na mente humana, ou seja, como são feitas as operações neurais responsáveis por armazenar uma grande quantidade de vocabulário e determinar como essas unidades lexicais se encadeiam. No nível pragmático, opera como *corpus* de investigação de fenômenos da linguagem humana e de validação de regras de encadeamento. Em resumo, a gramática proposta pelo estudioso, coleta e registra amostras do uso autêntico da língua em situações interacionais para validar e identificar fenômenos linguísticos não catalogados.

Propor um modelo de organização da língua é fundamental para a linguística computacional, uma vez que a efetividade de um aplicativo que interage com o usuário é calcada no pressuposto de que ambos (aplicativo e usuário) usem o mesmo código para comunicação. Por isso, o conceito de gramaticalidade introduzido nas pesquisas de Noam Chomsky em análise das línguas naturais é aplicado ao desenvolvimento de aplicativos que reconhecem ou sintetizam a fala (OTHERO; MENUZZI, 2005).



A noção de gramática usada na linguística computacional é a mesma apresentada há mais de meio século por Chomsky e sustentada por Cristóvão Silva (2003, p. 15) quando diz que “uma gramática deve explicitar os princípios e as características da língua analisada”. Em suma, para a linguística computacional, a gramática funciona como “um conjunto de regras que nos permitirá, juntamente com um determinado léxico, formar todas as sentenças gramaticais de uma dada língua” (OTHERO; MENUZZI, p. 51, 2005).

Esse conhecimento serve à linguística computacional como ponto de partida para a programação de um aplicativo qualquer de reconhecimento ou síntese de fala, pois se debruça sobre a forma de organização sintática das línguas naturais e opera com funcionamento quase matemático, o que justifica sua recorrência na concepção e desenvolvimento de recursos que envolvem Inteligência de Reconhecimento de Fala (SRI<sup>1</sup>).

Reiterando, a proposta desses aplicativos parte do princípio de que existe nas línguas uma forma lógica de estruturação e que, um sistema alimentado com informações referentes ao léxico e com a forma estrutural de uma determinada língua, pode operar atendendo às necessidades do usuário quanto à realização de tarefas simples, como efetuar uma ligação, realizar uma pesquisa online ou transcrever o que lhe foi ditado. Geralmente estas ações são efetuadas via dispositivos móveis, como *tablets*, celulares e computadores pessoais (MARTINS; BRASILIANO, 2012).

Entretanto, para que o funcionamento dessas ferramentas seja o mais eficaz possível, é preciso que o usuário fale o idioma em que o dispositivo está configurado. Muitos dos aplicativos de reconhecimento de fala disponíveis no mercado foram desenvolvidos para um público alvo que fala inglês. Isso poderia significar que grande parte dos usuários pudesse usar o aplicativo sem impedimentos, tendo em vista que a língua inglesa é falada por 1.500.000.000 de pessoas ao redor do mundo, em que apenas 375 milhões desse total são falantes nativos do idioma (STATISTA, 2016).

Se esse cálculo pode ser mantido em relação ao número de usuários de aplicativos de reconhecimento de fala, então se torna evidente a necessidade de aferir se a interação entre o usuário falante de inglês como língua

---

<sup>1</sup> *Speech Recognition Intelligence* (tradução minha).

estrangeira<sup>2</sup> e aplicativos de reconhecimento de fala é funcional, já que três a cada quatro falantes de inglês a usam como língua estrangeira (L2). Dentro desse parâmetro, tem-se que reconhecer o papel do Inglês como Língua Franca (ILF), haja vista que as comunicações entre indivíduos ocorrem independente de suas L1s. A ideia abarcada pelo termo ILF contempla o real usuário dos recursos de reconhecimento de fala uma vez que prevê a multiplicidade de perfis linguísticos que esses usuários podem ter.

## 1.1 JUSTIFICATIVA

As variações entre um modo de pronúncia e outro quando se aborda a produção de fala de um determinado idioma pode ser entendida como sotaque. Nas palavras de Derwing e Munro (2009, p.447), “o sotaque tem sido responsabilizado por diversas coisas. Ele tem sido visto como a causa pela falta de comunicação, e tem sido usado para camuflar racismo e outras formas de discriminação. Ele tem sido, inclusive, visto como um tipo de distúrbio que requiere remediação<sup>3</sup>”.

Mas, além de ser uma questão de sotaque, alguns aspectos da fala podem ter sutilezas mensuráveis que implicam diretamente na inteligibilidade da fala. É o caso da produção das consoantes oclusivas em posição de ataque silábico em sílabas tônicas por não falantes de inglês como língua materna, que pode gerar o desentendimento de várias palavras devido ao fato de existir um grande número de pares mínimos em português e inglês que contrastam as oclusivas surdas e sonoras em posição de *onset* absoluto em sílabas tônicas.

Em Português Brasileiro (PB), as consoantes /p, t, k/ são produzidas por seus falantes nativos com aspiração antes da soltura completa de ar e início da vibração das cordas vocais. Isso, em PB, pode não acarretar em falta de inteligibilidade e os seus usuários são capazes de diferenciar essas oclusivas surdas com aspiração das suas correspondentes sonoras (em que há vibração de cordas vocais antes da desobstrução da cavidade oral) /b, d, g/. Em outros

<sup>2</sup> O presente trabalho não faz diferença entre os termos “inglês como segunda língua” (ISL) e “inglês como língua estrangeira” (ILE).

<sup>3</sup> *Accent has been blamed for all sorts of things. It has been seen as the cause of miscommunication, and it has been used as a cover-up for racism and other kinds of discrimination. It has even been viewed as a kind of disorder that requires remediation.*(tradução minha)

idiomas, como o Inglês, esse tipo de ocorrência pode levar o ouvinte a perceber as oclusivas surdas como sonoras.

O Quadro 01 apresenta alguns exemplos dessa dualidade que é representada por “dois vocábulos que se distinguem apenas por um fonema” (CARDOSO, 2009, p.44).

<b>Consoante oclusiva<sup>4</sup></b>	<b>Vozeamento</b>	<b>Português</b>	<b>Inglês</b>
Bilabial	Surda	pata (/ˈpatə/)	pig (/pɪg/)
	Sonora	bata (/ˈbatə/)	big (/pɪg/)
Alveolar	Surda	tela (/ˈtɛlə/)	time (/taɪm/)
	Sonora	dela (/ˈdɛlə/)	dime (/daɪm/)
Velar	Surda	cola (/ˈkɔlə/)	class (/klæs/)
	Sonora	gola (/ˈgɔlə/)	glass (/glæs/)

QUADRO 01 – Pares mínimos das oclusivas surdas e sonoras em PB<sup>5</sup> e IA<sup>6</sup>.

Estudos como os de Schwartzhaupt e Alves (2014), Bent e Bradlow (2003) e Cruz (2006; 2008) discorrem sobre diferenças de produção e inteligibilidade, principalmente no que tange o desempenho oral em língua estrangeira, que dão suporte a novas pesquisas e enriquecem a produção de saber nesse campo de conhecimento.

Por outro lado, existem tipos de interação que não envolvem, necessariamente, dois falantes humanos e não foram contemplados pelas pesquisas supracitadas. É o que acontece, por exemplo, quando um indivíduo utiliza mecanismos tecnológicos para comunicação ou execução de alguma outra tarefa. Há alguns anos, esse tipo de tecnologia vem otimizando o tempo e facilitando a vida dos usuários, que se utilizam dos aprimoramentos informáticos para estabelecer relações mais rápidas e precisas com o mundo.

Quanto a esse aspecto, os estudos já feitos e as ferramentas tecnológicas desenvolvidas preocupam-se, sim, com as diferentes formas de produção de determinado idioma pelos usuários e, por consequência, operam para melhorar a percepção (e reconhecimento) da fala desses usuários. Aparelhos de telefonia móvel, por exemplo, carregam em sua configuração básica um dispositivo de reconhecimento de voz que apresenta características

<sup>4</sup> Conforme o ponto de articulação.

<sup>5</sup> Português Brasileiro.

<sup>6</sup> Inglês Americano.

humanoides e interage, dentro de certos limites, com o usuário. Esses recursos estão disponíveis em diversos idiomas incluindo o Português Brasileiro. Quanto ao inglês, os *softwares* geralmente carregam em suas opções de configuração variedades do idioma, dentre eles o Australiano, Canadense, Indiano, Neozelandês, Singapuriano, Britânico e o Americano.

Entretanto, apesar de considerar falantes de todas essas variedades de inglês, pesquisas realizadas indicam que o número de falantes de inglês como língua estrangeira supera o número de falantes nativos do idioma. Jenkins (2002), assim como Seidlhofer (*apud* Cruz, 2006), afirmam que 80% das interações em Inglês não incluem falantes nativos. Ao passo que o Inglês como Língua Franca (ILF) ganha espaço num contexto global, esse pode ser considerado um indicativo de que propostas de estudos com enfoque no falante de ILE tornem-se fundamentais para o desenvolvimento e aperfeiçoamento da tecnologia já existente no mercado.

Para o nosso conhecimento, não há registro, na academia brasileira, de investigações que se debrucem sobre a interação homem-máquina evidenciando os aspectos linguísticos que limitam o uso de recursos informáticos. Assim, uma vez que o uso de recursos tecnológicos tem facilitado o aprendizado de línguas e que o acesso à informação tem crescido e acelerado continuamente, a importância da proposta do presente trabalho se justifica na iniciativa em abrir um campo de investigação que leva em conta uma interação de ouvintes, falantes e tecnologia não contemplada até então, para o nosso conhecimento, especialmente estudos com aprendizes brasileiros de inglês.

## 1.2 OBJETIVOS

Levando em consideração as constatações discutidas e tendo em vista que investigações que tem como enfoque a inteligibilidade da interação usuário falante brasileiro de ILF e aplicativo de reconhecimento de fala humana não foram ainda conduzidas (para o nosso conhecimento), ambiciona-se, com esta pesquisa, dentre outras coisas, abrir caminho para que outras investigações com interesses semelhantes se debrucem sobre essa forma de interação relativamente recente.

Além disso, espera-se que os resultados obtidos possam melhorar a qualidade de programação dos aplicativos já existentes no mercado e que norteiem a concepção de novas ferramentas mais eficientes, possibilitando a mais usuários poderem desfrutar dos benefícios decorrentes do uso desse tipo de tecnologia.

Assim, as escolhas metodológicas desta pesquisa tiveram como questionamento norteador a seguinte inquirição: a produção das oclusivas surdas do inglês em posição de ataque silábico por brasileiros aprendizes de inglês serão inteligíveis a ferramenta de transcrição automática do YouTube?

Para fins metodológicos, estima-se responder à pergunta de pesquisa negando ou confirmando as seguintes hipóteses:

- a) As palavras-alvo produzidas pelos usuários brasileiros em que aparecem os segmentos /p, t, k/ em posição de ataque silábico serão menos inteligíveis que o segmento em si.
- b) Os segmentos /p, t, k/ em posição de ataque silábico serão mais inteligíveis em palavras-alvo familiares aos usuários brasileiros.

Para fins de comparação, os dados produzidos pelos brasileiros aprendizes de inglês também foram expostos a um grupo de ouvintes formado por 2 ouvintes nativos de inglês americano e 2 ouvintes brasileiros.

Os objetivos detalhados serão discutidos e mais bem justificados ao longo dessa dissertação, que está organizada em 5 capítulos, em que o primeiro tem teor introdutório. O CAPÍTULO 2 traz a fundamentação teórica na qual se apresenta e discute conceitos de inteligibilidade, se apresenta um resumo dos conhecimentos adquiridos no que cerne ao VOT e sua interface com a inteligibilidade, e se aborda os conceitos relacionados à tecnologia de reconhecimento de fala humana. Essa discussão permitirá que se encontrem as facetas de contato dessas três áreas e trará indícios de como elas dialogam.

O CAPÍTULO 3 descreve a metodologia usada na pilotagem e na coleta final dos dados para análise, bem como os participantes voluntários que se dispuseram a colaborar com essa investigação, a elaboração dos experimentos de coleta e a análise dos dados coletados. A descrição da metodologia também permitirá identificar limitações em que se deve atentar em futuras investigações.

O CAPÍTULO 4 traz os resultados dessa pesquisa, separados por parâmetros articulatórios. É neste capítulo que apresentamos também as análises estatísticas e os resultados dos experimentos conduzidos. O CAPÍTULO 5, por fim, apresenta as conclusões, limitações e aponta sugestões para trabalhos futuros relacionados à interface proposta nessa investigação.

## 2 FUNDAMENTAÇÃO TEÓRICA

As áreas da Linguística que se preocupam com a produção e percepção dos sons que formam as línguas são a Fonética e a Fonologia. Esses ramos linguísticos buscam desvendar e entender o sistema de organização, produção e percepção dos sons de uma língua desde os princípios biológicos até as interferências do ambiente em que são produzidos.

Apesar de semelhantes, até mesmo no nome, e trabalharem com o mesmo objeto de estudo (sob olhares e conceitos diferentes), as duas áreas distinguem-se por linhas bastante tênues. Santos e Souza (2011, p.9) difere os termos argumentando que “a fonética trabalha com os sons propriamente ditos, como eles são produzidos, percebidos e que aspectos físicos estão envolvidos em sua produção. A fonologia opera com a função e organização desses sons em sistemas”.

A Fonética se preocupa em descrever como os sons da fala são produzidos, atentando para suas características articulatórias, perceptuais e acústicas, conforme Cristófaró Silva (2003, p. 23) detalha quando diz que a fonética é “a ciência que apresenta os métodos para a descrição, classificação e transcrição dos sons da fala, principalmente aqueles sons utilizados na linguagem humana”. Por sua vez, a Fonologia abrange um campo mais abstrato e trata da organização dos sons de acordo com padrões acústicos e articulatórios, além de organizar os sons de acordo com suas funções, e não de acordo com suas diversas características fonéticas (CRISTÓFARO SILVA, 2003). Segundo Cristófaró Silva (2003, p. 17) “A fonologia estabelece os princípios que regulam a estrutura sonora das línguas, caracterizando as sequências de sons permitidas e excluídas na língua em questão”. A aquisição de uma língua natural se dá no nível fonológico primeiramente, seguida do aprendizado da leitura e da escrita.

Embora exista uma correlação entre a representação verbal gráfica e oral, as letras do alfabeto nem sempre correspondem aos sons representados ortograficamente por elas, e isso explica porque muitos indivíduos em fase de alfabetização confundem-se quando têm que aplicar seus conhecimentos fônicos à ortografia. Santos e Souza (2011) exemplificam essa afirmação dizendo que “as crianças, que estão aprendendo a grafia, escrevem *xícara* com

*ch*, *exame* com *z*, *malha* com *lia*. Isso ocorre porque, na ortografia, um som não necessariamente corresponde a uma letra” (SANTOS; SOUZA, 2011, p. 25).

Esse tipo de confusão acontece não somente na transferência da linguagem oral para a linguagem escrita quando um mesmo indivíduo está em processo de aquisição de um idioma, mas ela é percebida, também, em outros contextos, como quando da interação oral de dois indivíduos com L1<sup>7</sup> distintas (pois há transferência de padrões de L1 para L2 e vice-versa), quando, em estágio inicial de aquisição de L2<sup>8</sup>, um indivíduo não adquiriu ainda todos os traços específicos da produção de um fonema em língua estrangeira, ou quando o sistema de reconhecimento de voz de determinado dispositivo não contempla as especificidades da fala de seu usuário, como o é o que se pretende investigar nessa pesquisa.

## 2.1 INTELIGIBILIDADE

Quando se trata de língua estrangeira, muito dificilmente alguém irá encontrar um professor que diga para seus alunos que o modelo de pronúncia ideal é aquele que se confunde com o de um nativo (KENWORTHY, 1987). Pelo contrário, no contexto globalizado, parece razoável que não se ignore as diversas interferências que recaem sobre a aprendizagem de línguas estrangeiras, como a mescla de padrões linguísticos entre a língua materna e a língua estrangeira em aquisição, o contexto de aprendizagem e até mesmo os objetivos comunicacionais pertencentes a cada aprendiz. Por isso, a inteligibilidade parece ser o objetivo mais razoável de se ter como meta no processo de aquisição de língua estrangeira. Mas, o que se entende por inteligibilidade não é consenso entre os pesquisadores (CRUZ, 2007) que, além de apresentarem definições diversas, desmembram o termo em outros níveis, detalhando-os cada um a sua maneira.

Assim, o conceito de inteligibilidade tem sido debatido desde a década de 1950, quando o teórico Catford (1950 *apud* CRUZ, 2007) debruçou-se a definir este construto. Para ele, o conceito de inteligibilidade recai sobre a compreensão do ouvinte a respeito das palavras enunciadas pelo falante.

---

<sup>7</sup> Língua materna.

<sup>8</sup> Língua estrangeira.



Catford contrapõe o conceito de inteligibilidade ao de eficácia, que se refere à resposta dada pelo ouvinte a uma enunciação do falante. Esses dois termos constituem, portanto, os dois níveis, ou dimensões (como ele definiu), inter-relacionados envolvidos no uso de um idioma. Essa conexão faz com que um discurso inteligível possa ser ineficaz, uma vez que o ouvinte pode não responder de acordo com o enunciado emitido. Por outro lado, um discurso ininteligível, dificilmente será eficaz (CRUZ, 2007).

Em 1985, Smith e Nelson corroboraram as definições de inteligibilidade apresentando uma nova perspectiva. Para eles, o uso de uma determinada língua pode se desdobrar em três níveis (ao invés de dois, como proposto por Catford em 1950) – inteligibilidade, compreensibilidade e interpretabilidade. O primeiro termo refere-se ao reconhecimento de palavras ou enunciados. O segundo, ao significado atribuído à palavra ou ao enunciado emitido, enquanto o conceito de interpretabilidade recai sobre o sentido que se pode depreender de tal palavra ou enunciado.

Essa nova organização proposta dá a entender que o termo inteligibilidade se restringe à identificação de um determinado enunciado como parte de um dado idioma. Ou seja, se um enunciado qualquer é emitido e o interlocutor ouvinte percebe este como um grupo de palavras pertencentes à língua em uso na situação comunicativa, o interlocutor falante pode ser descrito como inteligível.

Ainda na mesma década, Kenworthy (1987, p.13) definiu inteligibilidade de uma forma bem concisa, simplificando esta dimensão em termos de “ser entendido por seu ouvinte interlocutor num dado tempo e numa dada situação<sup>9</sup>”. A autora ainda ressalta que todas essas abordagens têm enfoque no papel que o ouvinte ocupa na situação interacional. Ou seja, se o falante troca um som por outro e isso faz com que o ouvinte ouça uma palavra ou frase diferente do que foi a intenção do locutor emitir, pode-se dizer que o locutor é ininteligível (KENWORTHY, 1987).

Embora pareçam abrangentes, as definições apresentadas usam como referência a comunicação com um falante nativo de determinado idioma. Quando aplicado ao contexto de língua estrangeira, outros fatores interferem

---

<sup>9</sup> *Intelligibility is being understood by a listener at a given time in a given situation.*

na inteligibilidade, como Becker (2013, p.50) enfatiza ao dizer que “com a expansão do uso da língua inglesa, a ideia de inteligibilidade também sofreu uma expansão considerável, fazendo com que a própria noção do que é estrangeiro fosse colocada em cheque”.

Reconhecer outros contextos linguísticos em que o inglês, por exemplo, é usado como ferramenta de comunicação é fundamental ao estudo de inteligibilidade, principalmente porque há fatores específicos dessa interação que restringem a inteligibilidade e outros nem tanto. Um exemplo clássico é a presença, ou ausência, de sotaque. Uma vez que se entende que o propósito de falar um idioma é ser compreendido por um interlocutor (nativo ou não), questões como o sotaque podem não interferir na inteligibilidade e, por isso, esta ser uma dimensão irrelevante à comunicação (BECKER, 2013).

Ainda, num contexto global em que a quantidade de falantes de inglês como língua estrangeira é maior que o número de falantes nativos desse idioma, talvez os diversos graus de sotaque não precisem ser uma preocupação que justifique a matrícula em um curso de redução de sotaque, por exemplo, mas devem ser levados em consideração se influenciarem diretamente na inteligibilidade, no sentido de modificar completamente a sequência fonológica que representa um determinado vocábulo, fugindo, assim, da amplitude variacional esperada para um som específico.

Para estudos que se debruçam sobre a inteligibilidade, torna-se importante saber que a língua materna tem interferência sobre a aquisição de língua estrangeira e vice-versa. Reconhece-se que a transferência que ocorre entre L1 e L2 pode acarretar possíveis restrições à inteligibilidade na interação falante estrangeiro/falante nativo ou mesmo falante estrangeiro/falante estrangeiro que utilizem uma língua comum (adicional) para a comunicação. Isto é, sabe-se que por mais desenvolvida que seja a competência oral de dado falante de língua estrangeira, ele pode não ser compreendido em decorrência de incompatibilidades nos alfabetos fonéticos (ou inventário sonoro) das línguas materna e estrangeira (GILNER; MORALES, 2010) e das interferências que um acarreta no outro, dentre outros fatores que também podem ser responsáveis pela falta de inteligibilidade, como os que se encontram no nível suprasegmental.

Isso acontece porque, muitas vezes, sons distintivos em um idioma não causam distinção em outro sistema linguístico (SHWARTZHAUPT; ALVES, 2014) e, uma vez que o falante pode não sentir necessidade de diferenciá-los, a comunicação pode ficar comprometida. Isto é, quando essa transferência de padrões de uma língua para outra se dá no nível segmental da língua, um único traço fonético pode ser a exclusiva pista acústica que diferencia duas palavras, como em *bin* (cesto, lata) e *pin* (senha, alfinete) (GILNER; MORALES, 2010). Neste exemplo, os diferentes segmentos consonantais que iniciam as palavras fazem com que sejam atribuídas cargas semânticas diferenciadas aos vocábulos.

Na comparação, os traços distintivos que estabelecem a diferença semântica das palavras em destaque diferem apenas em sonoridade (/p/ é surdo e /b/ é sonoro) e este traço se mostra essencial para a inteligibilidade de um enunciado. Com diferenças tão pontuais como a descrita (diferença no início da vibração das cordas vocais), a tarefa do aprendiz de línguas estrangeiras se intensifica em ter que comparar e contrastar os alfabetos fonéticos de duas (ou mais) línguas estabelecendo limites claros entre a produção de um e de outro.

Por isso, durante a fase de aprendizado de uma segunda língua, um indivíduo apresenta tendências a adaptar os conhecimentos linguísticos adquiridos no exercício de sua língua materna à língua estrangeira em estágio de aquisição (SCHWARTZHAUPT; ALVES, 2014). Segundo os autores, “uma vez que os elementos fônicos de L1 e L2 dos aprendizes interagem em um espaço fonológico comum, esses aprendizes tendem a não perceber que características acústicas fazem parte dos sistemas da L1 e da L2 em questão” (SCHWARTZHAUPT; ALVES, 2014, p. 53). Jenkins (2000) ainda enfatiza que dificuldades em adquirir competência de percepção e produção de determinado fonema em língua estrangeira são atribuídas à mencionada transferência de padrões fonológicos de L1 para L2, que pode acarretar interferências danosas à interpretação, ao reconhecimento e, por conseguinte, à inteligibilidade.

Mesmo assim, a interação indivíduo/indivíduo pode não ser completamente comprometida, uma vez que existe a negociação de sentidos e a possibilidade de uso de formas alternativas de se fazer entender, como a linguagem gestual, o uso de sinônimos e até mesmo a definição de termos. Em

outros contextos, entretanto, a qualidade da competência oral do falante torna-se crucial para garantir a inteligibilidade da comunicação, como se evidencia na interação indivíduo/máquina, negligenciada até então.

Mesmo com um arcabouço denso de pesquisas que discutem inteligibilidade (KENWORTHY, 1987; CRISTÓFARO SILVA, 2003; CRUZ, 2007; SMITH e NELSON, 1985; CATFORD, 1950; BECKER, 2013; GILNER e MORALES, 2010; SCHWARTZHAUPT e ALVES, 2014), tem-se que destacar que, como não há consenso entre os pesquisadores quanto à abrangência e aos domínios do termo, há também uma carência de metodologias que avaliem esse construto e de pesquisas que apontem quais são os traços ou características que mais interferem na inteligibilidade (BECKER, 2013). Por ser um construto relativamente novo no que tange às análises linguísticas (entrou em voga a partir da década de 1950), determinar uma metodologia eficiente que avalie e aponte os caminhos para o entendimento da inteligibilidade é tarefa desafiadora ao linguista.

Um aspecto que atinge a acuidade de avaliação da inteligibilidade são as diversas interferências que variáveis podem causar nos resultados e invalidar os dados coletados. Se a produção da amostra for feita através da leitura de um texto pré-definido, por exemplo, deve-se atentar ao domínio que o participante tem desta habilidade (caso contrário se estará avaliando a capacidade de leitura do participante) e à influência do registro escrito na fala (BECKER, 2013).

Becker (2011), em pilotagem de experimentos, definiu que a leitura, em detrimento da fala espontânea, se prestaria à avaliação de inteligibilidade uma vez que há relativo controle de variáveis que se mostrariam fundamentais à análise do construto, como a velocidade de fala, o emprego de vocabulário simplificado, pausas no decorrer do discurso, entre outras estratégias. A autora ainda discute as desvantagens da fala espontânea para pesquisas que discutem inteligibilidade, apontando que este tipo de registro

não permite o controle de certas variáveis como a estrutura prosódica do enunciado em que se localiza um determinado som, alvo da pesquisa, e impede também quantificar os dados, já que é pouco provável que o mesmo som se repita sob as mesmas condições linguísticas, como por exemplo, o ambiente fonológico (BECKER, 2013, p. 61).

É notado que a escolha de uma metodologia em detrimento de outra apresenta vantagens e, também limitações e essa decisão depende amplamente do tipo de evidência linguística que se busca (MUNRO, 2008). Para pesquisas de viés quantitativo, como é o caso desta, a coleta controlada de dados através da leitura de um texto pré-determinado parece ser a forma mais apropriada de computar amostras que detenham informação linguística pertinentes ao que se pretende investigar. Conforme afirma Xu (2010, p.6), “de um modo geral, as características associadas à fala de laboratório estão mais relacionadas ao propósito da pesquisa que à fala de laboratório em sentido amplo. Quando queremos entender vogais, consoantes e tonicidade, temos que estar aptos a controlar a variação desses aspectos da fala enquanto mantemos outros constantes<sup>10</sup>”. Foi esta a estratégia usada nesta pesquisa e que será mais bem detalhada no CAPÍTULO 3, que descreve a metodologia empregada da coleta de amostras.

Tratando-se de inteligibilidade de língua estrangeira, os detalhes linguísticos que compõem a fala de um indivíduo não têm o mesmo peso para a determinação do que é inteligível ou não em um enunciado qualquer. Algumas nuances são mais determinantes que outras e isso se dá diferentemente em cada língua, como se mostra quando há transferências de padrões linguísticos da L1 para a L2. No caso do inglês como língua franca, evidenciado nesta pesquisa, e mais especificamente no caso do que foi convencionalmente chamada por Jenkins (2000) de *Lingua Franca Core* (LFC), a transferência linguística se dá principalmente no nível fonológico.

O termo *língua franca* faz referência a “uma língua de contato usada entre povos que não compartilham uma primeira língua, e é comumente entendida como querendo significar uma segunda (ou subsequente) língua de seus falantes<sup>11</sup>” (JENKINS, 2007 *apud* BECKER, 2013, p. 31). Em princípio, o termo foi usado para referir-se exclusivamente à comunicação entre indivíduos

---

<sup>10</sup> *In general, the characteristics attached to lab speech are related to the purpose of the study rather than to lab speech in general. When we want to understand vowels, consonants and tones, we have to be able to control the variation of these aspects of speech while keeping other aspects constant* (tradução minha).

<sup>11</sup> *In essence, a lingua franca is a contact language used among people who do not share a first language, and is commonly understood to mean a second (or subsequent) language of its speaker* (tradução de Becker, 2013).

que não compartilhassem o inglês como língua materna, mas expandiu-se a descrever interações que incluem interlocutores dos mais diversos perfis.

Jenkins (2000), em seus estudos, identificou cinco áreas que seriam as responsáveis pela preservação da inteligibilidade na interação entre indivíduos que usam o inglês como língua estrangeira. São elas:

- o inventário de consoantes;
- o agrupamento de consoantes (*clusters*);
- a distinção na duração de vogais (longas x curtas);
- o acento tônico - *nuclear stress*;
- os requisitos fonéticos (e.g. aspiração após /p/ , /t/, /k/).

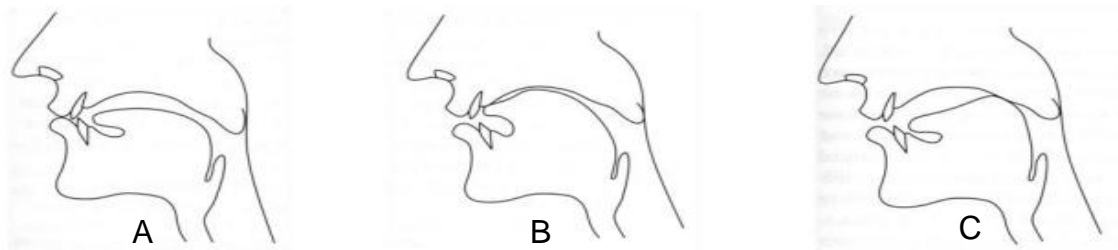
Os que nos interessam para fins metodológicos são os requisitos fonéticos, no que se estende à produção das oclusivas surdas.

## 2.2 AS OCLUSIVAS SURDAS

Mesmo dentro de um mesmo idioma, encontram-se variações nos sons que compõem o alfabeto fonético da língua. Algumas dessas diferenças são bastante evidentes e não é necessário conhecimento linguístico avançado para que sejam notadas, como se pode perceber na enunciação de palavras terminadas em S em língua portuguesa. Na região sul do Brasil, por exemplo, o último fonema da palavra *lápis* será mais frequentemente produzido como uma consoante fricativa alveolar surda: [ˈla.pɪs], enquanto no Rio de Janeiro, se um carioca reproduzir a mesma palavra, notaremos a palatalização da fricativa em coda: [ˈla.piʃ] (LAMPRECHT, 2004).

Quanto à aquisição dos sons, os segmentos oclusivos orais são os que se destacam como os primeiros a serem adquiridos no processo de aprendizagem de uma língua. As oclusivas bilabiais (/p, b/) são as que aparecem em primeiro lugar na linha de aquisição, seguidas pelos pares /t, d/ e, mais tardiamente, /k, g/. Reiterando, “pesquisas como as de Hernandorena (1990) e Lamprecht (2004) afirmam que plosivas e nasais são os primeiros segmentos consonantais a serem adquiridos, estando estabelecidos antes dos dois anos de idade” (LAMPRECHT, 2004, p. 75).

FIGURA 01 – Articulações bilabial (A), alveolar (B) e velar (C).



FONTE: ROACH, 2009 *apud* SOUZA, 2012.

As oclusivas orais são produzidas através de uma obstrução total da cavidade oral de tal modo que a passagem de ar é completamente impedida, seguida da abertura desse bloqueio, fazendo com que a corrente egressiva de ar pulmonar contida na cavidade oral seja solta. De acordo com o que Laver (1994, p.205) diz sobre as condições mínimas para a caracterização das oclusivas orais,

Um segmento oclusivo é definido como um segmento cuja fase medial é caracterizada por uma constrição oral completa ocasionada pela aproximação do articulador ativo ao passivo. Isso evita que o ar escape pela boca e, se uma constrição velar é mantida de modo que o ar também não pode escapar pela cavidade nasal, sob condições de pressão de ar egressiva, a pressão intra-oral [...] aumentará.<sup>12</sup>

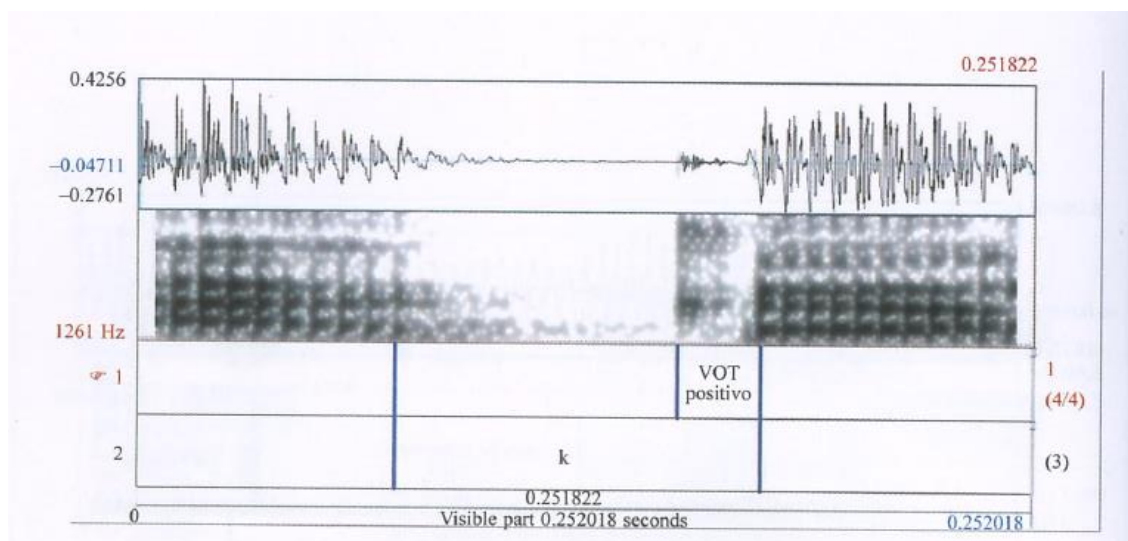
A obstrução descrita acontece acima da região laríngea enquanto o vozeamento acontece na laringe, que abriga as cordas vocais. Se as cordas vocais encontram-se unidas, mas maleáveis, e a sua vibração acontece antes da desobstrução do tubo acústico, dizemos que as realizações são sonoras. Se, ao contrário, as cordas vocais estão afastadas e a soltura de ar precede a vibração das cordas vocais, temos as realizações chamadas surdas.

<sup>12</sup> A stop segment is defined as a segment whose medial phase is characterized by a stricture of complete oral closure made by the active articulator against the passive articulator. This prevents the escape of air from the mouth, and if a simultaneous velic closure is maintained so that the air cannot escape through the nasal cavity either, then under conditions of positive egressive pressure, the intra-oral pressure [...] will rise (tradução minha).

De acordo com Lisker e Abramson (1964), a realização fonética das consoantes oclusivas pode variar de língua para língua em termos de uma relativa diferença no tempo de soltura da obstrução e o início da vibração das cordas vocais, o que é convencionalmente chamado de *Voice Onset Time* (VOT). Além disso, os autores dizem que se o vozeamento começa antes da soltura da obstrução, os valores de VOT serão negativos ([*-VOT*]); mas, se a vibração das cordas vocais acontecer depois da desobstrução do aparelho fonador, os valores de VOT serão positivos ([*+VOT*]).

As figuras 2 e 3 ilustram, em espectrograma e onda, o momento que é definido como VOT em uma consoante oclusiva surda e uma sonora em Português Brasileiro.

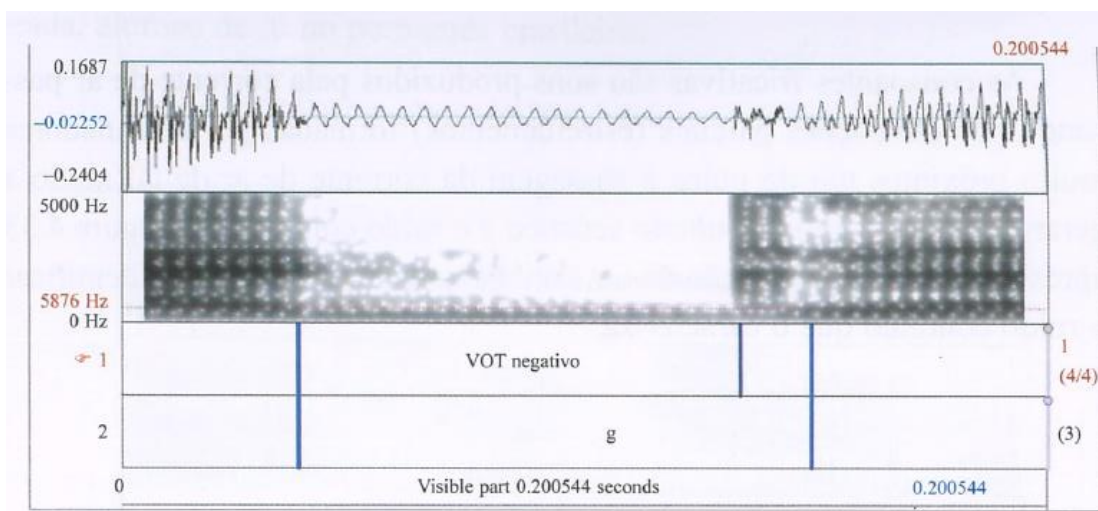
FIGURA 02 – Forma de onda e espectrograma da produção da consoante [k] entre vogais do PB.



FONTE: BARBOSA, 2015, p.174.



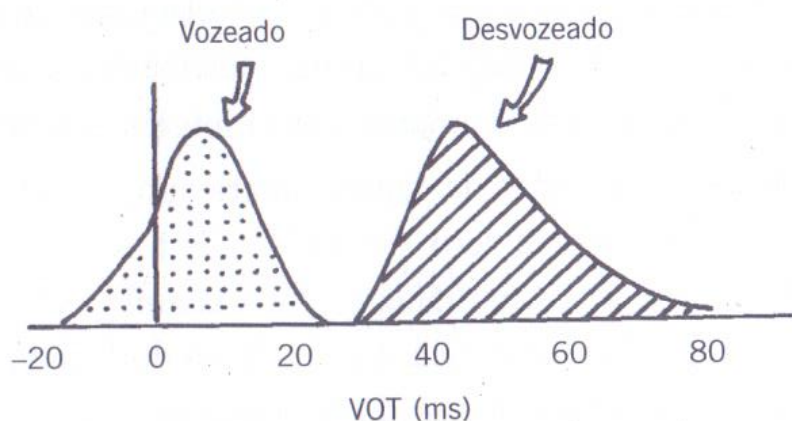
FIGURA 03 – Forma de onda e espectrograma da produção da consoante [g] entre vogais do PB.



FONTE: BARBOSA, 2015, p.175.

Kent e Read (2015) indicam que os valores de VOT carregam informações suficientes para a determinação do vozeamento e do ponto de articulação das oclusivas. Ainda que não exista um valor fixo que possa ser aplicado a todos os falantes de todas as línguas em todos os contextos fonéticos, os autores apresentam um esquema de distribuição desses valores baseados nas investigações de Azou (2000, *apud* KENT; READ, 2015, p.248), ilustrado pela Figura 4.

FIGURA 04 – Distribuição dos valores de VOT para as oclusivas vozeadas e desvozeadas.



FONTE: KENT; READ, 2015, p.248.

Essa variação encontrada na produção de oclusivas em um mesmo ou em idiomas diferentes era atribuída a um nível baixo de implementação das regras fonológicas ou a fatores biomecânicos individuais, mas não constituía uma parte necessária da gramática de uma língua qualquer.

Essas afirmações se confirmaram equivocadas à medida que os estudos relacionados à Fonética e Fonologia avançaram. Pesquisas que trazem dados de VOT em diferentes línguas e contextos mostraram que existem regras que são somente aplicáveis a determinadas línguas (CHO; LADEFOGED, 1999) e, por isso, a aquisição dessas características pode não compor traço distintivo para determinado grupo de aprendizes. Dessa forma, para os aprendizes de língua estrangeira cuja língua materna não contrasta as sutis variações na produção de oclusivas, neste caso, a aquisição dos traços de sonoridade torna-se irrelevante, no sentido em que o aprendiz não percebe a necessidade de desenvolver tal distinção.

É o que acontece, como já descrito, com a aspiração que precede a desobstrução da cavidade oral na produção das oclusivas surdas /p, t, k/ por falantes brasileiros de Português Brasileiro. Como, em PB, essa aspiração não faz com que o interlocutor, também brasileiro falante de PB, perceba a consoante surda como sonora, no aprendizado de Inglês (língua em que essa aspiração faz com que as oclusivas /p, t, k/ sejam percebidas como sonoras), esses falantes brasileiros não serão capazes de perceber a aspiração como um traço distintivo (pois ele não o é em sua língua materna) e, portanto, aplicar a distinção na L2.

Zimmer e Alves (2015, p.157) lembram que aprender os aspectos fonético-fonológicos de um dado idioma é um processo dinâmico e complexo, pois “Muitas variáveis, agindo em conjunto, mostram-se fundamentais para o entendimento desse complexo quadro”. Quando na aquisição de um segundo idioma, o aprendiz precisa desenvolver habilidades para perceber e empregar diferenças fonológicas do sistema linguístico da língua estrangeira (ZIMMER; ALVES, 2015).

Apesar de as consoantes oclusivas surdas em língua inglesa (/p, t, k/) serem um dos primeiros segmentos a serem adquiridos pelos falantes de língua estrangeira, os aprendizes podem não perceber a diferença de pronúncia entre a sua própria e a de um falante nativo. A aquisição do VOT

pode ser entendida como um desses desafios enfrentados pelos aprendizes de ILE. Enquanto em inglês as diferenças de VOT, e a aspiração pré vozeamento, são entendidas como distintivas para consoantes oclusivas surdas e sonoras, essa mesma variação em duração de VOT parece não ser imprescindível para o reconhecimento dessas consoantes aos aprendizes brasileiros (SCHWARTZHAUPT; ALVES; FONTES, 2013).

Ainda que em um nível fonético, o conhecimento das alterações que os sistemas linguísticos sofrem contribui para o desenvolvimento e domínio do idioma em aquisição. Entretanto, Schwartzhaupt e Alves (2014, p.53) atentam para o fato de que “essa transferência remete à noção de que pistas acústicas têm relevância diferente em dois sistemas, e o aprendiz é desafiado a perceber essa diferença, indo contra o seu conhecimento de L1”.

Zimmer e Alves (2015, p.158) afirmam que o falante de PB confia em outras pistas acústicas para a diferenciação das oclusivas surdas e sonoras, como “a intensidade da explosão do segmento plosivo e o valor de F0 da vogal seguinte”. Isso faz com que os aprendizes brasileiros possam ter a inteligibilidade de sua produção oral afetada, ou seja, ouvintes nativos de inglês podem encontrar dificuldades para estabelecer a diferença entre oclusivas surdas e sonoras na fala de um aprendiz de ILE brasileiro e, por isso, não estabelecer comunicação fluida.

Em contrapartida, Cho e Ladefoged (1999) afirmam que os falantes não produzem valores de VOT diferentes deliberadamente e o VOT é apenas “a inevitável consequência de movimentos fisiológicos e forças aerodinâmicas<sup>13</sup>” (CHO; LADEFOGED, 1999, p.225). Mas, se informação suficiente sobre os gestos envolvidos na produção oral for transmitido ao falante, os valores de VOT em diferentes línguas podem ser aprendidos e transferidos de uma língua à outra. Isso não significa dizer que as trocas e interferências obrigatoriamente acontecerão, mas elas podem ser tão determinadas quanto a mudança na frequência dos formantes que acontecem com gestos vocálicos específicos (CHO; LADEFOGED, 1999).

Lisker e Abramson (1964) apontam para as variações conhecidas de VOT relacionadas ao ponto de articulação das oclusivas (surdas e sonoras),

---

<sup>13</sup> *the inevitable consequence of the physiological movements and the aerodynamic forces* (tradução minha).

também observadas por Cho e Ladefoged (1999) e Schwartzhaupt e Alves (2014). As principais descobertas são: (1) quanto mais posterior for a obstrução, mais longo será o VOT; (2) quanto mais extensa for a área de contato dos articuladores ativo e passivo, mais longo será o VOT; e (3) quanto mais rápido for o movimento do articulador ativo, mais curto será o VOT.

Segundo esses parâmetros, esta pesquisa empenha-se em aferir se os valores de VOT influenciam na inteligibilidade da interação usuário-máquina, uma vez que essa interferência já foi constatada na interação entre indivíduos humanos. Nesse sentido, buscou-se aprofundar o olhar sobre os segmentos que poderiam oferecer maiores chances de reconhecimento, ou seja, os que se adequavam às descobertas apresentadas em Cho e Ladefoged (1999) e Schwartzhaupt e Alves (2014).

## 2.3 RECONHECIMENTO DE FALA

A tecnologia de reconhecimento de fala tem ocupado um lugar de destaque quando o assunto é inovação. Esse tipo de recurso tem sido alvo de pesquisa e desenvolvimento desde a década de 1950, com os primeiros estudos em Inteligência Artificial (AI<sup>14</sup>), mas foi somente algumas décadas mais tarde que o uso da voz como mecanismo de operacionalização de aplicativos obteve uma melhora significativa e sistemas de reconhecimentos puderam ser criados (MARTINS; BRASILIANO, 2012).

Nas últimas décadas, artigos tecnológicos, principalmente vinculados à comunicação, ganharam destaque em função desses recursos de reconhecimento de fala que são intrínsecos aos aparelhos. A comunicação entre usuários, que havia sido reduzida à troca de mensagens de texto, tem sido facilitada com o recurso de gravação de áudio, por exemplo. Isso pode ser visto como um indicativo de que a linguagem oral seja mais eficiente e acelera o processo de comunicação entre interlocutores, pois carrega informações linguísticas que são mais dificilmente percebidas na linguagem gráfica. Nuances de tonicidade e prosódia, por exemplo, muitas vezes não são capturadas na transcrição do que se diz ou quer dizer.

---

<sup>14</sup> Do inglês, *Artificial Intelligence*.

Ainda, a velocidade em que se fala é mais acelerada que aquela em que se digita, e isso pode implicar demora e falha na comunicação entre usuários de um mesmo recurso tecnológico. Isso se reflete também nos canais de comunicação em que o usuário interage com uma máquina (computador, *smartphone*, *tablet*, etc.). Para promover maior agilidade no processamento de operações simples, o usuário é convidado a emitir comandos de voz à máquina que, por sua vez, responderá de acordo com uma programação específica.

Alguns recursos representativos da interação dessa natureza são os comandos de voz em ligações de SAC (Serviço de Atendimento ao Consumidor), acionamento de dispositivos em aparelhos celulares (realizar ligação, conectar à rede de internet, programar alarme, etc.) e conversão de texto falado para escrito (transcrição automática).

Para que essas funcionalidades sejam possíveis, é necessário que o dispositivo tecnológico em uso esteja programado para operacionalizar de maneira a compreender o *input* oral de comando e realizar o que foi solicitado. No caso de aplicativos de reconhecimento de fala com a função de transcrição automática, o aplicativo recebe o *input* oral na forma de fala do usuário, analisa e decodifica as informações sonoras conforme o banco de dados em que foi configurado e converte a informação sonora em texto (ADAMI, 2010).

Ferramentas com função de reconhecimento de fala podem operar de duas maneiras, com ou sem treinamento. O reconhecedor com treinamento realiza uma seção prévia ao seu uso efetivo de reconhecimento de padrões de fala. Ou seja, ele pede que o usuário leia trechos de texto para identificar como é a produção dos fonemas por aquele usuário. Isso, em tese, facilitaria o processo de reconhecimento, já que, além dos dados já armazenados, o *software* teria esses dados da fala do usuário como mais um item na matriz de cruzamento. No caso da ferramenta de transcrição do YouTube, o mecanismo de reconhecimento não trabalha com treinamento, e usa somente o seu banco de dados para determinar a transcrição das amostras sonoras a que é exposto.

Para isso, os aplicativos de reconhecimento de fala, em sua maioria, operam com um sistema que é carregado com amostras de fala humana e uma gramática simplificada que atua como facilitadora dos processos de probabilidade que determinam qual item lexical selecionar. Esse tipo de processo é conhecido como modelo acústico, em que o sistema opera para

determinar a probabilidade de reconhecimento de uma determinada sequência sonora em função do que foi efetivamente pronunciado. Apesar de parecer um procedimento simples, o processo de decodificação torna-se complexo devido à natureza do *input*, pois o reconhecimento de fala funciona a partir do cruzamento de informações que determinam o item vocabular contido no banco de dados do aplicativo correspondente à evidência acústica fornecida pelo usuário (ADAMI, 2010).

Dentre os modelos acústicos já desenvolvidos com aplicabilidade em SRI<sup>15</sup>, o que apresenta maior eficiência na tarefa de reconhecimento de sinais de fala é o Modelo Escondido de Markov (HMM<sup>16</sup>), descrito entre as décadas de 1960 e 1970 (OLIVEIRA; MORITA, 2005). Conforme Andrade (2000, p.38) descreve, “uma cadeia de Markov é um conjunto finito de elementos, formando uma máquina de estados. Nesta máquina de estados, as transições entre os estados não são governadas por regras determinísticas, mas por probabilidades de transição entre eles”, ou seja, o HMM é um sistema autômato que gera uma sequência de símbolos discretos a partir de um determinado estímulo, mas não ocupa-se em expor ao usuário os caminhos que percorreu para determinar o produto final dos processos desempenhados.

Para esclarecer, Andrade (2000, p. 39), dá o exemplo de um senhor que vai à feira, a pedido de alguém, comprar os itens descritos em uma lista. Neste caso, houve a interação de um interlocutor (que emite a solicitação) com alguém que realiza determinada atividade e lhe apresenta o produto final (unidades discretas). Tais unidades discretas podem ser entendidas como os produtos que este senhor traz para casa, e que o solicitante tem a chance de avaliar a conformidade do que foi trazido com o que lhe foi demandado. Partindo dessa lógica, Andrade (2000, p. 39) diz que

não está explícito nas frutas qual foi o caminho percorrido pelo consumidor através das finitas barracas e o que motivou esse caminho. Aspectos como distâncias, preços, quantidades e qualidades das frutas influenciaram no trajeto, e estes aspectos podem possuir características estocásticas.

---

<sup>15</sup> Do inglês, *Speech Recognition Intelligence*.

<sup>16</sup> Do inglês, *Hidden Markov Model*.

No HMM, portanto, o que fica visível ao usuário é somente o *output*, enquanto os trajetos e combinações feitas através dos processos probabilísticos não são evidenciados. Aplicado a um sistema de reconhecimento e transcrição de fala, o usuário tem acesso à representação escrita do que o aplicativo reconheceu dentre as possibilidades encontradas após o cruzamento de informações acústicas e gramaticais, mas não lhe é fornecido informações sobre as operações realizadas para que isso fosse possível. Segundo Oliveira e Morita (2005, p.2), “a maior vantagem do HMM situa-se na sua natureza probabilística, apropriada para sinais corrompidos por ruídos tal como a fala [...]”, ou seja, este modelo opera bem com a inconstância de sinais que variam no tempo, como o são os sinais acústicos.

Mesmo assim, se os *softwares* armazenassem em sua memória operacional sequências inteiras de sons referentes a dado idioma, as chances de identificarem uma palavra seriam quase nulas, uma vez que cada ocorrência de dado som é praticamente única, haja vista que “a fala espontânea não permite o controle certas variáveis como a estrutura prosódica do enunciado em que se localiza um determinado som [...] e impede também quantificar os dados, já que é pouco provável que o mesmo som se repita sob as mesmas condições linguísticas, como por exemplo, o ambiente fonológico” (BECKER, 2013, p.61). Ainda, se um sistema de reconhecimento de fala tivesse um sistema de programação que comportasse sons, vocábulos e gramática inteiros em sua memória, sua funcionalidade seria diminuída pela carga de informações contidas, que freiam a velocidade de operação. Ou seja, aplicativos com sistema de dados muito extensos, mesmo os que funcionam online, são mais lentos e comprometem a eficácia de uso de seus recursos.

Uma das coisas que justifica a recorrência de uso de HMM na configuração e aplicativos de reconhecimento de fala é a possibilidade que esse sistema oferece de manipular e analisar variações em atributos fragmentados. Andrade (1999, p.23) esclarece que

Os sinais acústicos da fala podem ser convertidos em uma sequência de números que represente a variação de amplitude no decorrer do tempo. A sequência pode ser subdividida em quadros ou janelas que se superpõem, abrangendo amostras vizinhas. E destas janelas pode-se extrair atributos da voz que sejam mais representativos do que puramente uma sequência de amplitudes.

Por isso, a maioria dos sistemas opera com fragmentos de sons, ou seja, cada elemento sonoro que irá compor o banco de dados base é capturado e fragmentado em unidades ainda menores que a extensão de sua imagem espectrográfica. Apenas algumas dessas unidades são selecionadas e passam a compor o banco de dados dos sistemas de reconhecimento de fala que, com sua estrutura de programação estatística, seleciona a sequência de palavras que seja mais provável de ser produzida, observando-se os dados sonoros emitidos pelo usuário. Isto é, o aplicativo reage ao *input* sonoro e, num jogo de probabilidades e combinação, tenta determinar a qual item do seu sistema, salvo em sua memória operacional, a sequência sonora emitida corresponde (ADAMI, 2010).

Mas reconhecer palavras isoladas, como os comandos dados a um aparelho de telefonia móvel, é menos complexo que reconhecer as palavras encadeadas de uma sentença. Quando inseridas em uma estrutura frasal, as palavras sequenciais ultrapassam as suas barreiras de início e fim e se sobrepõem, causando o que é convencionalmente chamado de coarticulação. Para que esse não seja um empecilho ao uso dessa tecnologia, os Modelos Escondidos de Markov operam com modelos de palavras extraídas de diferentes frases numa tentativa de contemplar os diferentes contextos fonológicos que podem ocorrer na fala corrida.

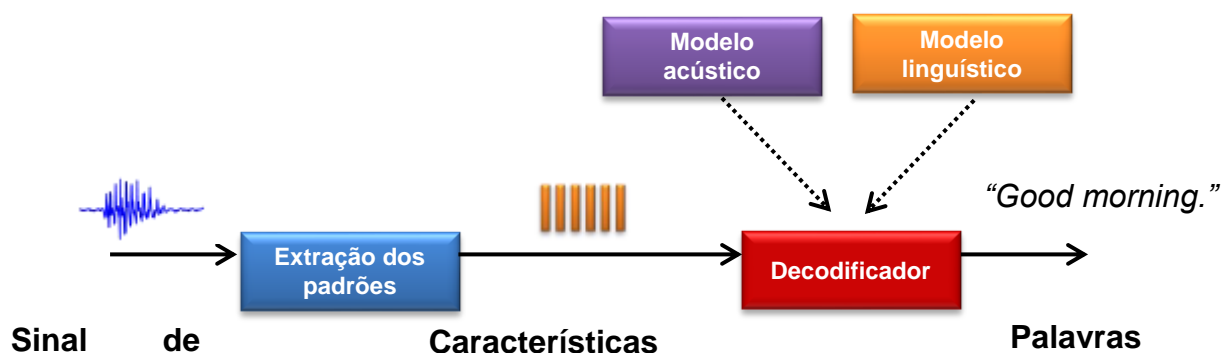
Ainda, numa tentativa de deixar mais eficiente esse modelo de reconhecimento, alguns aplicativos armazenam, também, uma gramática referente às normas que regem a sequência de palavras em uma frase de determinada língua (inspirada na gramática chomskiana), que serve como modelo linguístico. Na prática, “O uso de uma gramática que estabeleça uma ordem na sequência das palavras pode facilitar a busca do [aplicativo] pela sentença correta [...], evitando de pesquisar todas as palavras modeladas” (ANDRADE, 1999, p. 27). Ou seja, o aplicativo trabalha com um sistema de probabilidade que indica que categorias de palavras se encaixam melhor na sequência sintática das frases. Por exemplo, numa frase de estrutura sujeito-verbo-objeto, as palavras que mais provavelmente ocupem a posição de sujeito são pronomes e substantivos (podendo, é claro serem acompanhados de outros termos acessórios). Nesse sentido, ao invés de realizar uma busca por



todo o seu arcabouço lexical, o sistema do aplicativo vai restringir a busca a essas duas classes gramaticais, acelerando e tornando, muitas vezes, o sistema mais eficiente.

Essa relação é também descrita por Louzada (2010, p. 11) quando explica que “em sistemas de reconhecimento de fala que usam comparação de padrões, as características do sinal de fala desconhecido são comparadas com o padrão previamente armazenado e o padrão mais próximo do sinal de entrada é escolhido”. A Figura 05 ilustra os principais componentes de um sistema de reconhecimento de fala. Nele, pode-se perceber o trajeto que o sinal sonoro percorre e a interação das informações armazenadas com as fornecidas até o ponto em que é decodificado e convertido em palavras escritas.

FIGURA 05 – Arquitetura de um sistema de reconhecimento de fala



FONTE: ADAMI, 2010, p.5.

No esquema da Figura 05, a evidência acústica é emitida pelo usuário, no primeiro estágio, e capturada pelo sistema de reconhecimento. A próxima etapa do processo consiste em analisar e fragmentar o sinal de fala e, então, estabelecer os padrões sonoros que, já no decodificador, serão comparados ao modelo acústico armazenado no sistema da máquina e cruzados com o modelo linguístico que restringirá a gama de vocábulos que podem ser compatíveis à sequência frasal enunciada, culminando na transcrição da fala.

A escolha da ferramenta que serve como interlocutor da interação proposta nos experimentos descritos no Capítulo 3 dessa dissertação (transcritor automático do YouTube) foi baseada nessa função, pois se trata de um dispositivo informático que independe de treinamento, ou seja, o aplicativo

não precisa passar por uma seção prévia de reconhecimento de padrões de fala do usuário, principalmente por estar disponível online. Entende-se que a sua disponibilidade gratuita na rede faz com que sua popularidade e, conseqüentemente, emprego atinja um público expressivo (mais de um bilhão de usuários até agosto de 2017<sup>17</sup>) e, por isso, sua configuração tende a ser mais elaborada, tendendo a atender uma ampla diversidade linguística.

---

<sup>17</sup> Dados disponíveis em: <<https://www.YouTube.com/intl/pt-BR/yt/about/press/>> Acesso em: 07 Ago. 2017.

### 3 PROCEDIMENTOS METODOLÓGICOS

Durante os estágios iniciais da pesquisa, várias opções metodológicas vieram à tona principalmente por se tratar de uma investigação pioneira na academia (para o nosso conhecimento). Na falta de referências que apontassem para um caminho similar ao que se decidiu percorrer, buscou-se amparo na publicação de Derwing, Munro e Carbonaro (2000) na qual descrevem um experimento que tem como enfoque “a eficácia de um sistema de reconhecimento automático de fala para a produção oral de Inglês como Segunda Língua<sup>18</sup>”.

Os participantes do experimento conduzido pelos pesquisadores (DERWING; MUNRO; CARBONARO, 2000) foram organizados em dois grupos, ouvintes e falantes. O grupo de falantes (que forneceriam as amostras sonoras) foi composto de 30 indivíduos cuja língua materna era cantonês, espanhol ou inglês canadense (10 participantes para cada língua materna). O grupo dos ouvintes foi composto por 41 homens e mulheres, todos falantes de inglês canadense. O procedimento de coleta foi dividido em três etapas: “(a) treinamento do aplicativo de reconhecimento de fala (Dragon NaturallySpeaking Preferred, 1997) para reconhecer a língua dos participantes; (b) elaboração das amostras de fala em ILE e obtenção dos níveis de reconhecimento do aplicativo; e (c) fazer com que os ouvintes transcrevessem e avaliassem as amostras de fala<sup>19</sup>” (DERWING; MUNRO; CARBONARO, 2000, p. 595).

Para a primeira etapa descrita no texto de Derwing, Munro e Carbonaro (2000), treinamento do sistema de reconhecimento, os participantes do grupo de falantes foram orientados a ler três capítulos do livro “3001: The Final Odyssey”, de Clarke (1997). Na segunda fase, solicitou-se a esses mesmos participantes que procedessem com a leitura de 60 frases em um microfone atrelado ao computador, para que o aplicativo pudesse fazer o reconhecimento. Na terceira fase, os pesquisadores criaram um *corpus* com duas amostras de

<sup>18</sup> “effectiveness of a popular ASR package for ESL speech” (tradução minha).

<sup>19</sup> “(a) training the ASR Software (Dragon NaturallySpeaking Preferred, 1997) to recognize the speakers' language, (b) creating ESL speech samples and obtaining recognition scores from the software, and (c) having the listeners transcribe and rate the speech samples” (tradução minha).

frases produzidas por cada participante (gerando um total de 60 sentenças) para que se pudesse expor aos participantes ouvintes de forma randomizada (DERWING; MUNRO; CARBONARO, 2000, p. 595-596).

Findas as etapas de coleta, prosseguiu-se à análise dos dados coletados. Para computar o nível de inteligibilidade, Derwing, Munro e Carbonaro (2000) contabilizaram a quantidade de frases reconhecidas pelo aplicativo e a porcentagem de palavras transcritas corretamente pelos participantes ouvintes. Os resultados apontaram que os sistemas de reconhecimento de fala não são tão precisos quanto o ouvinte humano principalmente no que tange o reconhecimento de sotaque. Quanto à produção em língua inglesa, o aplicativo testado (*Dragon NaturallySpeaking Preferred*, 1997) apresentou taxa de reconhecimento em concordância à anunciada pelo provedor.

Dos materiais disponíveis com temática de investigação similar, este foi o que mais se aproximou do que pretende esta pesquisa. Portanto, frente à escassez de pesquisas que pudessem elucidar as decisões metodológicas quanto à coleta de dados que se precisava, à falta de infraestrutura que possibilitasse uma coleta sem interferências externas (como ruídos nas amostras de áudio e acesso estável à internet, por exemplo) e à complexidade do que se pretendia investigar, conduziu-se experimentos piloto para validar os instrumentos elaborados.

A pesquisa em linguística, principalmente ao que cerne a inteligibilidade, apresenta limitações e depende de escolhas que irão determinar a qualidade e validade dos dados coletados (BECKER, 2013). Por isso, as escolhas metodológicas que determinaram a produção dos instrumentos de coleta desta pesquisa tiveram como questionamento norteador a seguinte inquisição: a produção das oclusivas surdas do inglês em posição de ataque silábico por brasileiros aprendizes de inglês serão inteligíveis a ferramenta de transcrição automática do YouTube?

Em desmembramento, estima-se responder à pergunta de pesquisa negando ou confirmando as seguintes hipóteses:

- a) as palavras-alvo produzidas pelos usuários brasileiros em que aparecem os segmentos /p, t, k/ em posição de ataque silábico serão menos inteligíveis que o segmento em si.

- b) Os segmentos /p, t, k/ em posição de ataque silábico serão mais inteligíveis em palavras-alvo familiares aos usuários brasileiros.

Para fins de comparação, os dados produzidos pelos participantes brasileiros também foram expostos a um grupo de ouvintes formado por 2 ouvintes nativos de inglês americano e 2 ouvintes brasileiros proficientes em inglês.

Para a investigação das hipóteses, foram elaborados experimentos que pudessem servir como ferramenta de coleta de dados para que se pudesse confirmar ou negar tais hipóteses. Portanto, em busca de suporte científico para conduzir a análise dos dados coletados, os seguintes instrumentos de coleta foram elaborados:

- 1 - Questionário Biográfico
- 2 - Questionário de Familiaridade
- 3 – Coleta de amostras de fala<sup>20</sup> (leitura de frases)
- 4 - Experimentos de Inteligibilidade

Todos os instrumentos foram elaborados durante o curso da disciplina Tópicos Avançados em Linguística Aplicada I, ministrada pelo Prof. Dr. Ronald Barry Martinez em 2015, cuja ementa previu o “desenvolvimento e uso de instrumentos de pesquisa [...] mais comuns e importantes na pesquisa de linguística aplicada, como questionários, testes de proficiência linguística, e sistemas de observação de aula, todos baseados em princípios científicos”<sup>21</sup>.

Os encontros semanais da disciplina também serviram como ambiente de validação de instrumentos. Todos passaram pela cuidadosa avaliação de meus colegas e professor, além do olhar sempre atendo de minha orientadora, para que lacunas pudessem ser preenchidas e os instrumentos melhorados. Tendo em foco a natureza dos dados que se pretendia coletar, a quantidade, a qualidade e a ordem de aplicação dos instrumentos foram determinadas.

Na sequência, faremos a descrição e apresentaremos justificativa para as escolhas metodológicas que motivaram a elaboração dos experimentos supracitados.

---

<sup>20</sup> A princípio, em pilotagem dos experimentos, objetivou-se a coleta das amostras de fala simultaneamente à interação com o aplicativo de reconhecimento de fala, e o experimento foi nomeado “experimento de interação”. Dados alguns fatores restritivos descritos mais adiante nesta dissertação, o procedimento de coleta foi alterado, assim como sua nomenclatura.

<sup>21</sup> A ementa está disponível no portal SIGA para alunos da UFPR.

### 3.1 METODOLOGIA DO ESTUDO PILOTO

Essas primeiras experiências foram feitas com colaboradores voluntários (incluindo-se a própria pesquisadora) que se submeteram à coleta de dados para que pudesse ser feita a validação dos instrumentos. Ao final, quatro pilotagens preliminares foram conduzidas até que se entendessem as limitações tecnológicas e de infraestrutura que poderiam vir a ser intervenientes. As amostras sonoras coletadas nas primeiras três coletas não serviram ao propósito da investigação, pois fatores como a interrupção do sinal estável de conexão com a rede de internet e ajustes no modelo de coleta tornaram inviáveis a manutenção e análise dos dados gerados. Essa etapa inicial, entretanto, teve grande validade para a construção dos instrumentos de coleta, especialmente os que não dependiam de conexão com a internet. A princípio pretendia-se que os aprendizes brasileiros lessem as frases ao aplicativo de reconhecimento que fazia a transcrição simultânea, e isso dependia de boa qualidade de conexão de internet. Os questionários biográficos e de familiaridade foram aplicados e se mostraram eficazes quanto ao que avaliavam durante a etapa de pilotagem.

A seguir, descrevem-se os instrumentos de coleta elaborados e a pilotagem feita com um grupo de voluntários que atendia às especificações biográficas que se buscava, assim como os ajustes que se mostraram necessários.

#### 3.1.1 População alvo e seleção dos sujeitos do estudo piloto

Para a pilotagem dos instrumentos delineados, foram selecionados alunos matriculados em curso de idiomas à data da coleta. Dos dez voluntários que se dispuseram a colaborar, sete concluíram a fase de pilotagem. Esses sete informantes obedeciam aos seguintes critérios:

- Nenhum dos sujeitos apresenta ou declarou imparidades na fala, audição ou visão;

- Os participantes tinham nível mínimo B1 de proficiência em língua inglesa<sup>22</sup>;
- Todos os participantes eram brasileiros;
- Todos os participantes foram alfabetizados em sua língua materna, o Português Brasileiro.

O Quadro 02 especifica as características biográficas desses participantes.

PARTICIPANTE	01	02	03	04	06	08	10
SEXO	Fem.	Fem.	Fem.	Fem.	Fem.	Masc.	Masc.
IDADE	23	15	30	19	36	38	23
NACIONALIDADE	Brasil	Brasil	Brasil	Brasil	Brasil	Brasil	Brasil
LÍNGUA MATERNA	PB	PB	PB	PB	PB	PB	PB
ALFABETIZAÇÃO	PB	PB	PB	PB	PB	PB	PB
IDADE INICIOU INSTRUÇÃO LI	21	12	12	8	33	99	12
VIVÊNCIA PAÍS ANGLOFALANTE	não	não	não	não	não	sim	sim
VIAGEM ANGLOFALANTE	não	não	sim	sim	não	não	sim
IMPARIDADE AUDITIVA	não	não	não	não	não	não	não
IMPARIDADE VISUAL	sim	sim	sim	não	sim	não	sim
IMPARIDADE FALA	não	não	não	não	não	não	não

QUADRO 02 – Características biográficas do grupo de voluntários para a pilotagem<sup>23</sup>.

A codificação atribuída aos sujeitos voluntários não obedece a uma ordem sequencial pois os indivíduos 05, 07 e 09 não puderam dar continuidade à sua participação.

Para interagir com o grupo de brasileiros, selecionou-se o aplicativo Voice Note II, disponível no endereço eletrônico <https://voicenote.in/>, que opera com a função “ditado” e oferece várias possibilidades de configuração de linguagem. Além disso, o aplicativo é compatível com diversos sistemas operacionais incluindo Mac OS, Windows, Linux e Chrome OS.

A busca pelos possíveis *softwares* foi feita através do site Google.com usando as palavras-chave *voice*, *recognition* e *software*<sup>24</sup>. A busca direcionou a

<sup>22</sup> Conforme o Quadro Comum Europeu.

<sup>23</sup> Em 2015.

pesquisa para a página de aplicativos do Google Chrome, navegador desenvolvido pela empresa Google. O software selecionado apresentou maior número de revisões dentre os disponíveis e atendeu às expectativas da pesquisadora quanto a sua funcionalidade.

O interesse pelo software descrito acentuou-se quando, em sua descrição de requerimentos de uso leu-se “Para usar o Chrome VoiceNote, você precisará de: [...] microfone configurado, pronúncia correta”<sup>25</sup>. Como a intenção deste projeto é identificar as características sonoras que interferem na inteligibilidade na interação sujeito-*software*, a definição do que se quis dizer com “pronúncia correta” pode ser questionável.

Além do grupo de participantes brasileiros e do aplicativo, com o objetivo de triangular os dados coletados, objetivava-se aplicar os instrumentos de coleta a participantes voluntários não brasileiros. Para tal, esses voluntários teriam que obedecer aos seguintes critérios:

- Não apresentar ou declarar imparidades na fala, audição ou visão;
- Ser nativo americano;
- Ter sido alfabetizado em língua inglesa;
- Não ter residência fixa ou temporária no Brasil.

As necessidades descritas se justificam em função do terceiro perfil de participante que serviria para a avaliação da inteligibilidade que se busca nesta pesquisa, o aplicativo. Dado que o aplicativo usado para conduzir o experimento de interação seria configurado em uma variante específica da língua inglesa (o inglês americano), precisava-se que esse grupo controle (formado pelos participantes não brasileiros) tivesse como língua materna o inglês americano (com limitação ao contato com outros idiomas).

Para a fase de pilotagem, todavia, não foram aplicados alguns dos experimentos, entre eles o de inteligibilidade, que expunha amostras de fala coletadas durante a pilotagem do experimento de interação à transcrição de participantes nativos americanos. Essa opção foi feita frente à falta de voluntários americanos que atendessem aos parâmetros biográficos estipulados.

---

<sup>24</sup> Do Inglês, voz, reconhecimento e programa.

<sup>25</sup> *To use the Chrome VoiceNote, you will need: [...] Configured microphone; Correctly pronunciation*” (tradução minha).



### 3.1.2 *Corpus* do estudo piloto

Para aferir a inteligibilidade de segmentos com diferentes VOTs produzidos pelos participantes brasileiros em interação com o aplicativo de reconhecimento de fala, foi preciso determinar em que níveis a inteligibilidade seria aferida (segmental, lexical, ou textual). Como dito anteriormente, os aplicativos de reconhecimento e síntese de fala são programados com uma gramática referente à língua de operação que utilizam. Como o aplicativo selecionado para a condução dos experimentos na fase de pilotagem tem como função primária transpor para linguagem escrita o estímulo sonoro, o nível de compreensão do aplicativo pode se mostrar mais eficiente quando as palavras-alvo são inseridas em contextos de nível textual, neste caso, em frases.

Isso significa que o ditado palavra a palavra poderia não oferecer o tipo de dado que se precisaria para conduzir a análise proposta. Mesmo assim, achamos válido que as palavras-alvo fossem empregadas em frases diferentes para verificar se essa alternância de ambiente fonológico constituía fator influenciável.

O corpus do estudo piloto, apresentado no Quadro 02, é constituído por três listas de palavras que evidenciam o contexto fonológico a que se pretendeu seguir:

- Segmento oclusivo bilabial surdo em início de sílaba e de palavra seguido de vogal anterior alta em palavras monossílabas ou dissílabas;
- Segmento oclusivo alveolar surdo em início de sílaba e de palavra seguido de vogal anterior alta em palavras monossílabas ou dissílabas;
- Segmento oclusivo velar surdo em início de sílaba e de palavra seguido de vogal anterior alta em palavras monossílabas ou dissílabas.

A escolha desses contextos determinou-se com base nos estudos de Yavas e Wildermuth (2006), que mostram que, para a investigação de VOT, o contexto descrito por /# p, t, k + i/ evidencia os valores VOT, pois a altura da vogal (anterior-alta ou posterior-alta) que segue as oclusivas surdas faz com que os valores de VOT sejam mais longos, facilitando sua identificação espectral.

Alves e Schwartzhaupt (2014) apontam, em corroboração, que a quantidade de sílabas da palavra-alvo não interfere nos valores de VOT para o contexto /# p, t, k + i/, ou seja, monossílabas e dissílabas não apresentam medidas contrastivas de VOT. Por isso, optamos por utilizar os vocábulos que atendessem ao contexto /# p + i/; /# t + i/; /# k + i/ e que fossem mais familiares aos participantes.

A familiaridade dos termos empregados era uma preocupação e uma variável que se pretendia controlar, haja vista que o desempenho oral do falante está diretamente relacionado ao quanto familiar lhe é o vocabulário (CRUZ, 2008). Frente a isso, previamente a esta escolha lexical, a escola permitiu que o material empregado nas aulas fosse analisado para verificar quais vocábulos aprendidos atendiam às necessidades fonológicas dessa investigação. Ao todo foram analisados 12 volumes do material destinados ao ensino de ISL ao público adolescente e adulto, e todas as palavras com a combinação [plosiva surda+vogal anterior alta] em posição de ataque silábico foram listadas. Os livros correspondiam às etapas já vencidas pelos estudantes participantes desta pesquisa e partiu-se do pressuposto de que as palavras selecionadas a partir dessa análise teriam um grau de familiaridade maior.

Como o material não trazia substantivos suficientes para equilibrar numericamente todas as categorias propostas (#p/t/k + vogal anterior alta) houve a necessidade de empregar o verbo *to keep*, modificando a posição que a palavra-alvo ocuparia na estrutura sintática proposta para a elaboração das frases veículo (sujeito + verbo + objeto + complemento). Além disso, os substantivos *keel*, *peak* e *teaser* também foram empregados mesmo não constando nos livros didáticos da escola, uma vez que se pretendia manter equilibrados os exemplares de palavras reais da língua de acordo com os critérios almejados (quantidade de sílabas e classe gramatical).

#/p/ + vogal anterior alta	#/k/ + vogal anterior alta	#/t/ + vogal anterior alta
picnic    ɪ    peach    i:	kitchen    ɪ    key    i:	ticket    ɪ    teeth    i:
period    ɪ    pizza    i:	kill    ɪ    kiwi    i:	till    ɪ    teen    i:
pink    ɪ    people    i:	kid    ɪ    keyboard    i:	tiptoe    ɪ    teaspoon    i:
pillow    ɪ    peanut    i:	kit    ɪ    keep    i:	teacher    i:
picture    ɪ    piece    i:	kitty    ɪ <b>keeper</b> i:	tea    i:
<b>Peak</b> i:	<b>keel</b> i:	team    i:

		<b>teaser</b> i:
--	--	------------------

QUADRO 03 – Unidades lexicais presentes no material da instituição de ensino de línguas que atendem ao contexto /p, t, k + vogal anterior alta/.

O Quadro 03 sintetiza os itens identificados no material didático da instituição e os que foram acrescentados (destacados em negrito), dadas as necessidades metodológicas da pesquisa. Cada um dos itens selecionados foi inserido em frases veículos que obedeciam à estrutura sintática sujeito + verbo + objeto + complemento. Além disso, tomou-se o cuidado de estabilizar o contexto fonológico em que cada vocábulo seria inscrito. Dessa maneira, optou-se pelo emprego de verbos terminados em consoantes oclusivas surdas, de modo que a consoante em análise pudesse ser mais bem evidenciada em análise espectral. O Quadro 04 demonstra a distribuição das palavras-alvo nas frases.

/p/ + /i/	<b>Monossílabas</b>	<i>peach</i>	<i>We ate peach yesterday.</i>
		<i>peace</i>	<i>We got peace like a river.</i>
		<i>peak</i>	<i>We walk peaks together.</i>
	<b>Dissílabas</b>	<i>pizza</i>	<i>We eat pizza for lunch.</i>
		<i>peanut</i>	<i>We ate peanuts yesterday.</i>
		<i>people</i>	<i>We hit people downstairs.</i>
/k/ + /i/	<b>Monossílabas</b>	<i>key</i>	<i>We got keys downstairs.</i>
		<i>keep</i>	<b><i>We keep coffee here.</i></b>
		<i>keel</i>	<i>We got keels to fix.</i>
	<b>Dissílabas</b>	<i>kiwi</i>	<i>We eat kiwi for lunch.</i>
		<i>keyboard</i>	<i>We got keyboards to fix.</i>
		<i>keeper</i>	<i>We help keepers downstairs.</i>
/t/ + /i/	<b>Monossílabas</b>	<i>teeth</i>	<i>We help teeth problems.</i>
		<i>teen</i>	<i>We help teens at school.</i>
		<i>team</i>	<i>We stop team fights.</i>
	<b>Dissílabas</b>	<i>teacher</i>	<i>We help teachers at school.</i>
		<i>teaspoon</i>	<i>We ship teaspoons for you.</i>
		<i>teaser</i>	<i>We talk teaser sentences.</i>

QUADRO 04 – Distribuição das palavras-alvo em frases veículo.

A única exceção foi marcada em negrito e corresponde ao emprego do verbo *keep*. Tal divergência de padrão foi mantida mesmo que fugisse à estrutura frasal determinada para que se pudesse alcançar o número

equilibrado de seis itens lexicais para cada contexto fonológico e para que se pudesse aferir o impacto que a alternância de função sintática causa na mensuração da inteligibilidade.

### 3.1.3 Gravações e coleta de dados do estudo piloto

Os dados coletados durante a fase de pilotagem não se estendem a todos os experimentos desenhados (Questionário Biográfico, Questionário de Familiaridade, Experimento de interação e Experimento de Inteligibilidade). Apenas foram gerados dados com os experimentos de familiaridade e de interação com o aplicativo, como se descreve a seguir. A coleta foi conduzida de agosto a outubro de 2015.

Para aferir a familiaridade dos itens lexicais escolhidos para compor o corpus do estudo piloto, criou-se um questionário a que os participantes brasileiros deveriam responder de forma objetiva a duas perguntas. Para cada item lexical apresentado aos participantes, foi-lhes perguntado:

(a) Você reconhece essa palavra?

(b) Você sabe o seu significado?

A cada uma das perguntas o participante poderia responder “sim” ou “não”, evitando gradiência nas respostas dadas. Quando a resposta foi negativa para qualquer das perguntas, apresentou-se ao participante um cartão ilustrado que se referia ao item lexical em questão, sem que a palavra lhe fosse efetivamente pronunciada pelo pesquisador, evitando, assim, interferência na pronúncia do participante. Dessa maneira, deu-se aos participantes a oportunidade de conhecer as palavras-alvo empregadas no experimento de interação imediatamente antes deste, evitando que a familiaridade se tornasse uma variável interveniente. O Apêndice E apresenta os 18 cartões que foram confeccionados para este experimento.

Aos termos investigados misturaram-se palavras-distratoras, que serviram como ponto de desfoque ao que o experimento de interação pretendia. As palavras distratoras escolhidas obedeciam, de certa forma, ao padrão das palavras-alvo, mas não foram empregadas nos experimentos de interação e inteligibilidade por não seguirem fielmente ao contexto fonológico determinado, como ilustra o Quadro 05.

CONTEXTO	POSIÇÃO	PALAVRA	VOGAL
#/p/ + vogal alta	Anterior	pencil	e
	Posterior	poor	ʊ
#/t/ + vogal alta	Anterior	table	e
	Posterior	tool	u:
#/k/ + vogal alta	Anterior	king	ɪ
	Posterior	cook	ʊ

QUADRO 05 – Palavras distratoras em função da consoante inicial.

Para determinar a quantidade de palavras distratoras que se misturaria às alvo obedeceu-se ao mesmo critério de Alves e Schwartzaupt (2014). Os autores escolheram o que correspondeu a cerca de 33% do número de palavras-alvo (duas por contexto fonológico). Dessa forma, o questionário de familiaridade continha um total de 24 itens que deveriam ser avaliados pelos participantes quanto ao seu reconhecimento como palavra pertencente ao inglês e quanto ao seu significado. Não houve problemas que causassem o descarte dos dados coletados com o questionário de familiaridade.

Para a coleta das amostras de fala e para a verificação da inteligibilidade imediata na interação dos participantes com o aplicativo de reconhecimento de fala, desenvolveu-se um experimento de interação que passou por diversas modificações até alcançar sua forma de pilotagem. Haja vista que se pretendia que cada participante lesse uma sequência diferente das mesmas frases, um documento de apresentação de *slides* foi elaborado e as frases foram embaralhadas manualmente conforme aleatorização feita previamente com o recurso online disponível no endereço *www.random.org*.

O instrumento foi validado por *member check* durante o curso da disciplina ministrada pelo Prof. Dr. Ron Martinez. A técnica de *member check*, conforme Cohen (2006), “acontece quando dados, categorias analíticas, interpretações e conclusões são testadas com os membros daqueles grupos com quem se coletou dados<sup>26</sup>”.

A gravação das amostras de fala foi feita em dispositivo periférico de captura de áudio<sup>27</sup>, a partir de um aplicativo padrão de gravação de áudio.

<sup>26</sup> “This is when data, analytic categories, interpretations and conclusions are tested with members of those groups from whom the data were originally obtained” (tradução minha).

<sup>27</sup> Ipod 5ª geração.

Após a gravação os dados foram armazenados e etiquetados. A captura de som contou com o auxílio de um headset com microfone acoplado<sup>28</sup>. No total, o estudo piloto contou com 756 amostras dados: 108 segmentos por sujeito (dois quais 36 para cada contexto fonológico e destes, 6 para cada palavra-alvo), multiplicados por 7 participantes.

As amostras de fala coletadas, entretanto, não ficaram isentas de ruídos devido à falta de espaço físico adequado para a coleta. A sala que nos foi disponibilizada não tinha tratamento acústico e as atividades pertinentes ao espaço que ocupávamos (escola) não foram interrompidas para a coleta. Isso, todavia, não foi entendido como um empecilho à coleta, mas fica registrado para referência futura.

#### 3.1.4 Limitações do piloto

A fase de pilotagem dos experimentos serviu para que ajustes pudessem ser feitos e para corroborar a validação feita por *member check*. Durante esse período, as limitações que impulsionaram ajustes e melhorias nos instrumentos foram registradas e se descrevem na sequência.

Percebeu-se que os computadores com apenas uma entrada para periférico do tipo microfone não permitem o uso do microfone para mais que um aplicativo por vez. Neste caso, pretendia-se empregar microfone para que fosse um ponto de entrada de áudio tanto para o programa de gravação de *input* sonoro quanto para assegurar a qualidade da interação com o aplicativo de reconhecimento de fala. Tendo de ser feita uma escolha, optou-se por empregar o microfone na interação com o sistema de reconhecimento de fala e um dispositivo móvel operou como gravador a fim de armazenar o mesmo *input* sonoro a que o aplicativo VoiceNote II reagia.

Identificou-se, também, outra limitação quando da aplicação do experimento de interação. Como a apresentação de *slides* com as frases veículo teria que ser feita no computador em uso pelo participante, não foi possível observar o comportamento do *software* de reconhecimento de fala durante a coleta. Apesar de não ter sido possível observar como o aplicativo

---

<sup>28</sup> Philips, modelo NL9206AD-4, ano 2016.

efetuava o reconhecimento e determinava os vocábulos a serem transcritos, essa limitação não se mostrou inteiramente intrusiva. Observou-se que, possivelmente, perceber a correspondência (ou não) entre a fala e a transcrição de sua própria fala poderia causar impacto negativo na produção oral dos participantes. Por isso, metodologicamente, entendeu-se que essa restrição informática não consistia em limitações à coleta de dados.

Ainda sobre as restrições identificadas, a qualidade da conexão de internet disponível e a ausência das condições de um estúdio de gravação se mostraram influentes na qualidade das amostras e no resultado do experimento em si. A primeira fez com que o aplicativo interrompesse a função de reconhecimento do aplicativo que se pretendia investigar e a segunda fez com que o áudio gravado em dispositivo móvel absorvesse muitos ruídos.

A dificuldade maior na coleta de dados com participantes americanos (como se pretendia em princípio) foi a falta deles. Devido às exigências biográficas específicas, encontrar participantes que atendessem às necessidades da proposta não foi tarefa simples. Dada essa dificuldade, o experimento de inteligibilidade não foi aplicado ao público-alvo desejado, mas foi elaborado para que se pudesse identificar a necessidade de ajustes metodológicos. A metodologia empregada para a coleta dos dados que foram efetivamente analisados para esta pesquisa será detalhada na seção seguinte.

### 3.2 METODOLOGIA DA COLETA DE DADOS DESTA PESQUISA

Para a coleta de dados que norteou as análises deste trabalho, lapidou-se os instrumentos pilotados para que pudessem corresponder às necessidades da proposta. Para retomar, a elaboração dos instrumentos teve início em 2015, durante a disciplina Tópicos Avançados em Linguística Aplicada I, ministrada pelo Prof. Dr. Ronald Barry Martinez e avançou conforme surgiram as necessidades descritas na seção anterior.

Cada instrumento elaborado (1 - Questionário Biográfico; 2 - Questionário de Familiaridade; 3 - Coleta de amostras de fala; 4 - Experimento de Inteligibilidade) também passou pela cuidadosa avaliação da Prof. Dra. Denise Cristina Kluge, orientadora desta investigação. Ao passo que novos conteúdos emergiam das leituras teóricas a respeito das três facetas em

interação nesta pesquisa (inteligibilidade, a produção de oclusivas e inteligência de reconhecimento de fala), ajustes foram aplicados aos instrumentos, conforme se descreve a seguir.

### 3.2.1 População alvo e seleção dos sujeitos

O perfil biográfico que se pretendia para a coleta dos dados atendeu ao descrito na seção 3.1.1 (População alvo e seleção dos sujeitos do estudo piloto). Entendeu-se que o perfil determinado atende às necessidades dessa investigação e pode-se classificar os participantes voluntários em dois grupos distintos, o que forneceriam as amostras de fala (grupo de falantes) e os que atestariam a inteligibilidades dessas amostras (grupo de ouvintes).

Na fase de determinação dos grupos de indivíduos que seriam convidados para compor o grupo de falantes, primeiro teve-se que descobrir onde poder-se-ia encontrar os participantes potenciais. Como o ensino de línguas estrangeiras evolui quanto a estratégias e métodos de ensino ao longo do tempo, nem todas as abordagens cumprem com as especificidades das necessidades que esta pesquisa apresenta. Para garantir que nossos participantes tivessem um tipo de instrução que possibilitasse a prática oral de uso da língua em situações comunicativas, optou-se por participantes matriculados em uma escola de idiomas que estimulasse a produção oral em suas práticas pedagógicas.

Por isso, após a investigação dos modelos metodológicos adotados por diversas unidades de ensino de língua estrangeira na cidade de Curitiba, Paraná, optou-se envolver participantes estudantes matriculados em um curso de idiomas que tem como base metodológica a teoria de aquisição de segunda língua o modelo behaviorista-estrutural. Esta proposta metodológica encara o aprendizado de línguas como uma formação de hábitos e tem como princípios fundamentais o pressuposto de que “a aprendizagem acontece através da repetição de estímulos; os reforços positivos e negativos têm influência fundamental para a formação dos hábitos desejados; [e] a aprendizagem ocorre melhor se as atividades forem graduadas” (PAIVA, 2014, p.13). O modelo proporciona uma estrutura metodológica de aulas que garantem ao aluno horas de instrução e prática da oralidade.



Além disso, a disponibilidade oferecida pelas instituições para que os alunos, professores e espaços fossem acessados teve um impacto favorável à coleta. Portanto, além de atender as necessidades técnicas fundamentais, os professores e a coordenação concordaram em ceder espaço para coleta de dados durante o horário de aula dos estudantes, o que fez com que os participantes se sentissem melhor acomodados. O Apêndice A traz um exemplar dos termos de consentimento que foram preenchidos e assinados pelos participantes voluntários e pela coordenação das escolas colaboradoras (à época de pilotagem e à época de coleta de dados desta pesquisa).

Após essa etapa de acordos, os participantes foram convidados a responder ao questionário biográfico para que o perfil do grupo de falantes pudesse ser traçado. O objetivo do questionário era o de determinar o perfil linguístico dos sujeitos e garantir que eles formassem um grupo homogêneo de voluntários participantes da pesquisa. Das características almejadas, buscava-se um grupo que não tivesse passado mais de 12 meses consecutivos em país de língua inglesa, que tivesse a língua portuguesa como língua materna, que não tivesse nenhum comprometimento auditivo e que tivesse sido instruído na variedade americana da língua inglesa.

Um exemplar do questionário biográfico aplicado aos grupos de falantes e ouvintes encontra-se disponível no Apêndice B desta dissertação. Entretanto, o preenchimento do questionário não foi feito sob supervisão da pesquisadora, e algumas dúvidas surgiram aos participantes, uma vez que o instrumento lhes foi entregue para preenchimento domiciliar. Apesar de terem sido instruídos a contatar a pesquisadora caso tivessem dúvidas quanto ao preenchimento, nenhum dos sujeitos o fez e, portanto, responderam ao questionário da maneira que acharam que seria mais adequada. Isso, todavia, não se mostrou como um impedimento para pesquisa e nem qualificou o descarte dos dados produzidos pelos participantes. De um modo geral, a questão referente às aulas de pronúncia foi respondida com a carga horária correspondente ao curso de idiomas em que estavam matriculados os respondentes.

Por isso, além de terem respondido ao questionário biográfico, alguns dos participantes foram entrevistados para que algumas das respostas apresentadas no questionário pudessem ser mais bem esclarecidas. O roteiro de entrevista seguido foi disponibilizado no Apêndice C.

Ao todo, então, os participantes do grupo de falantes somaram quatorze, dos quais sete responderam aos experimentos piloto (dois do sexo masculino e cinco do sexo feminino) e outros sete aos instrumentos finais (dos quais seis do sexo feminino e um do sexo masculino), obedecendo aos critérios de coleta determinados na pilotagem. O Quadro 06 retoma as características biográficas dos participantes da coleta piloto e descreve as dos outros sete voluntários que incorporaram o grupo de falantes. Para esta pesquisa, entretanto, as amostras de fala de apenas quatro participantes foram submetidas à transcrição da ferramenta do YouTube (sujeitos identificados pela marca 01, 02, 03 e 10), em **negrito** no Quadro 6). Isso se deu porque os dados coletados com os outros voluntários não foram reconhecidos pela ferramenta, que não ofereceu a possibilidade de transcrição automática.

SUJEITO	SEXO	IDADE	NACIONALIDADE	LÍNGUA MATERNA	LÍNGUA ALFABETIZAÇÃO	IDADE INICIOU INSTRUÇÃO LI	VIVÊNCIA PAÍS ANGLOFALANTE	VIAGEM PAÍS ANGLOFALANTE	IMPARIDADE AUDITIVA	IMPARIDADE VISUAL	IMPARIDADE FALA
01	fem.	25	Brasil	PB	PB	21	não	não	não	sim	não
02	fem.	17	Brasil	PB	PB	12	não	não	não	sim	não
03	fem.	32	Brasil	PB	PB	12	não	sim	não	sim	não
04	fem.	21	Brasil	PB	PB	8	não	sim	não	não	não
05	fem.	23	Brasil	PB	PB	11	sim	sim	não	não	não
06	fem.	38	Brasil	PB	PB	33	não	não	não	sim	não
07	fem.	23	Brasil	PB	PB	12	não	não	não	sim	não
08	masc.	40	Brasil	PB	PB	99	sim	não	não	não	não
09	fem.	27	Brasil	PB	PB	18	não	não	não	não	não
<b>10</b>	<b>masc.</b>	<b>25</b>	<b>Brasil</b>	<b>PB</b>	<b>PB</b>	<b>12</b>	<b>sim</b>	<b>sim</b>	<b>não</b>	<b>sim</b>	<b>não</b>
11	fem.	21	Brasil	PB	PB	19	não	não	não	sim	não

12	fem.	37	Brasil	PB	PB	26	sim	sim	não	sim	não
13	fem.	19	Brasil	PB	PB	14	não	não	não	sim	não
14	fem.	28	Brasil	PB	PB	13	sim	sim	não	não	não
15	fem.	28	Brasil	PB	PB	16	sim	sim	não	não	não
16	fem.	30	Brasil	PB	PB	7	não	não	não	sim	não
17	masc.	24	Brasil	PB	PB	7	não	não	não	não	não

QUADRO 06 – Características biográficas do grupo de falantes<sup>29</sup>.

Incluindo-se os 17 sujeitos, a média de idade dos participantes do grupo de falantes é 27 anos (em 2017), com uma variação de 17 a 40 anos. Como houve a restrição de reconhecimento dos arquivos submetidos à transcrição automática, os dados simplificados referentes aos sujeitos 01, 02, 03 e 10 são expostos no Quadro 07.

Sujeito	Sexo	Idade
01	feminino	25
02	feminino	17
03	feminino	32
10	masculino	25
<b>Média de idade</b>		<b>25</b>

QUADRO 07 – Características biográficas do grupo de falantes cujas amostras de fala foram submetidas à transcrição automática do YouTube.

Lembra-se que, para pesquisas que se debruçam sobre aspectos da fala humana, o gênero dos participantes pode ser significativo, uma vez que aspectos relacionados à voz podem ser afetados por essa variável. Pepiot (2013, p.2) retoma que “diferenças entre vozes feminina e masculina [...] não somente se referem à acústica (frequência fundamental, frequências de ressonância, etc.) e medidas perceptuais, mas também à anatomia e a

<sup>29</sup> As linhas destacadas em cor referem-se aos sujeitos que responderam ao questionário biográfico mas não procederam com a coleta de amostras de fala.

fisiologia (diferenças entre órgãos vocais), sociologia e até filosofia (construção da identidade de gênero, inato *versus* comportamento aprendido)<sup>30</sup>”.

Apesar de Pepiot (2013) apontar em investigação conduzida com falantes monolíngues de Frances e de Inglês (dois homens e duas mulheres para cada grupo) que os valores médios de VOT são significativamente mais longos para falantes mulheres que para falantes homens, pesquisas que evidenciam as implicações do gênero dos sujeitos nos valores de VOT mostraram que as diferenças, quando as há, não são significativas e tendem a ser relacionadas a outros fatores (WHITESIDE, HENRY, DOBBIN, 2004; MADHU SUDARSHAN REDDY et al., 2014). Dessa maneira, não percebeu-se a disparidade no número de homens e mulheres que compuseram o grupo de falantes como uma variável interveniente.

Para aferição de inteligibilidade sentiu-se necessário, também, a participação de um grupo de ouvintes falantes de inglês com nível de proficiência superior ao do grupo de falantes. Como resultado de opções metodológicas feitas anteriormente (variedade alvo da instituição de ensino e base de programação do aplicativo, por exemplo), precisávamos que o critério de formação do grupo de ouvintes obedecesse a alguns requisitos. O critério norteador baseou-se no atual contexto de Inglês como Língua Franca, em que a inteligibilidade alcança parâmetros diferentes, dado a interação com sujeitos de diferentes L1. O Quadro 08 evidencia a riqueza linguística dos sujeitos que compõem o grupo de ouvintes.

SUJEITO	SEXO	IDADE	NACIONALIDADE	LÍNGUA MATERNA	LÍNGUA ALFABETIZAÇÃO	PAÍS DE RESIDÊNCIA <sup>31</sup>	IMPARIDADE AUDITIVA	IMPARIDADE VISUAL	IMPARIDADE FALA
01.1	fem.	43	Brasil	PB	PB	Brasil	não	sim	não
02.1	fem.	38	Brasil	PB	PB	Brasil	não	sim	não

<sup>30</sup> *Differences between female and male voices [...] not only refer to acoustic (fundamental frequency, resonant frequencies, etc.) and perceptual measurements, but also to anatomy and physiology (differences in the vocal organs), sociology and even philosophy (construction of gender identity, innate versus learned behavior)* (tradução minha).

<sup>31</sup> Em 2017.

<b>03.1</b>	fem.	24	EUA	IA	IA	Brasil	não	sim	não
<b>04.1</b>	masc.	30	EUA	IA	IA	EUA	não	não	não

QUADRO 08 – Características biográficas do grupo de ouvintes.

A média de idade do grupo de ouvintes, conforme se evidencia, é de 34 anos, variando de 24 a 43 anos. O grupo foi composto, então, por duas brasileiras e dois americanos, um com residência nos Estados Unidos e uma com residência no Brasil. À época da coleta dos dados, a participante 03.1 estava morando em Curitiba, Paraná, há 5 meses. O sujeito 02.1 é nascido no Brasil, mas esteve morando na França por 12 meses entre 2015 e 2016. Dos participantes, este é o único caso em que o ouvinte falava mais de dois idiomas (incluindo a sua L1). Dos quatro participantes do grupo de ouvintes, os participantes 01.1, 02.1 e 03.1 trabalham com ensino de línguas e produção de material didático para o ensino de línguas. O participante 04.1 trabalha em empresa estatal americana, responsável pela emissão de carteiras de motorista e de identificação.

### 3.2.2 *Corpus* de coleta

Para a coleta de amostras de fala com os participantes do grupo de falantes, mantiveram-se as escolhas feitas à época de pilotagem, conforme apresentado na seção 3.1.3. As palavras-alvo e as frases se mostraram eficientes ao que se pretendia e a metodologia na escolha do *corpus* de coleta não foi alterada.

### 3.2.3 Gravação e coleta de dados

Para a coleta das amostras de fala que serviriam de base para aferir a inteligibilidade do objeto de investigação dessa pesquisa foram elaborados experimentos de três naturezas distintas, um de familiaridade, um de produção e um de inteligibilidade. Os testes foram aplicados aos participantes dos grupos de falantes e ouvintes em momentos e em sequência diferentes. A justificativa para essa decisão metodológica se pautava na necessidade de evitar

interferência no padrão de fala de ambos os grupos e garantir maior acuidade da verificação de inteligibilidade. O grupo de falantes respondeu aos experimentos de familiaridade e de produção, nesta ordem, enquanto o grupo de ouvintes foi submetido a avaliação de amostras de fala no experimento de inteligibilidade e, posteriormente, respondeu ao questionário de familiaridade.

Os participantes brasileiros aprendizes de inglês do grupo de falantes foram os primeiros a serem submetidos à coleta de dados. Após terem lido e assinado o termo de consentimento livre e esclarecido e respondido ao questionário biográfico, cada participante do grupo, individualmente, foi encaminhado, à uma sala silenciosa, haja vista a falta de estúdio com tratamento acústico nas instituições onde as amostras de fala foram coletadas. Já acomodados, a pesquisadora explicou aos informantes os procedimentos para coleta de dados. Percebeu-se necessário esclarecer como seria feita a coleta de dados e reiterar que o objetivo da pesquisa não era fazer qualquer julgamento sobre a qualidade de pronúncia do aprendiz, mas se o inglês que eles enunciavam, e que era compreensível para seus professores e demais interlocutores, era, também, inteligível ao aplicativo de reconhecimento de fala. As instruções ao grupo de falantes foram dadas em sua língua materna, neste caso, o Português Brasileiro.

Instruídos os participantes, aplicou-se o questionário de familiaridade do *corpus*. Este questionário foi elaborado para controlar a interferência que uma possível falta de familiaridade lexical pudesse implicar na qualidade da produção das palavras-alvo durante a coleta de amostras de fala. Assim, as palavras-alvo discriminadas anteriormente somadas a duas outras palavras distratoras para cada grupo fonológico investigado foram listadas uma a uma e seguidas de dois questionamentos objetivos, cujas respostas poderiam ser “sim” ou “não”.

Para cada um dos vocábulos listados, então, os participantes responderam se conheciam a palavra listada e se conheciam o seu significado. Quando a resposta a qualquer das perguntas foi não, entregou-se ao voluntário um cartão ilustrado contendo informações que poderiam auxiliar na compreensão do vocábulo. No cartão incluiu-se uma imagem ilustrativa do vocábulo, uma definição em inglês e um exemplo do termo empregado em sentença, conforme o Apêndice E. Quando o informante, mesmo com o

material de suporte apresentado, não entendia o significado do vocábulo, o mesmo era traduzido para o português pela pesquisadora. A fim de não influenciar a pronúncia dos informantes, em momento algum as palavras-alvo ou distratoras foram enunciadas pelos pesquisadores.

O Gráfico 01 que segue demonstra a familiaridade para cada um dos vocábulos listados no experimento de familiaridade e leva em consideração tanto o grupo de falantes inicial (com 14 indivíduos) quanto o grupo que teve suas amostras de fala transcritas pela ferramenta de transcrição do YouTube.

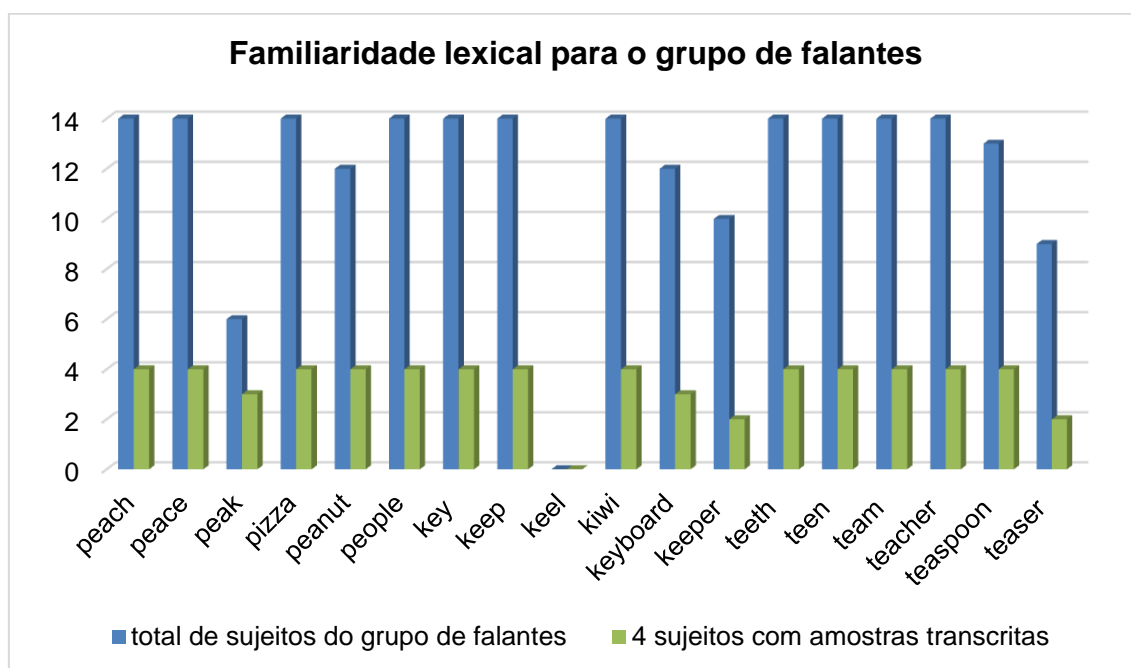


GRÁFICO 01 – familiaridade das palavras-alvo para o grupo de falantes.

Conforme ilustrado, as barras em verde referem-se à familiaridade de cada vocábulo-alvo para os quatro falantes que tiveram suas amostras de fala transcritas pelo YouTube e pelo grupo de ouvintes. Optou-se por fazer essa diferenciação para que se pudesse perceber com mais clareza as implicações da familiaridade lexical para a transcrição das amostras, de modo a identificar possíveis interferências relacionadas à pronúncia dos quatro falantes para o reconhecimento de fala.

Finda a etapa de aferição da familiaridade dos termos usados nos demais experimentos, os participantes foram acomodados em frente a um computador e explicamos a eles como o aplicativo que iriam usar funciona para a efetiva coleta de amostras de fala. Conforme mencionado anteriormente, em

função da baixa qualidade de conexão com a rede de internet, optou-se por realizar a coleta de dados sem a interação simultânea com uma ferramenta *online* de transcrição de linguagem oral para a escrita. Por isso, o computador empregado serviu ao propósito de apresentar em lâminas as frases veículo aleatorizadas para que cada participante pudesse lê-las em voz alta, como foi feito durante a pilotagem do experimento. Ressaltamos aos participantes a necessidade de uso de microfone<sup>32</sup> para a coleta das amostras de fala em função do armazenamento de informações com qualidade sonora que permitisse análise espectrográfica e reiteramos a garantia de anonimato.

Em seguida, os participantes foram orientados oralmente sobre os procedimentos específicos do experimento de interação. Como reforço, essas mesmas informações foram dispostas para leitura pré-coleta nas primeiras lâminas na apresentação de *slides* criada para o experimento, conforme se descreve no Apêndice F. Verificamos a compreensão dos procedimentos de coleta com os participantes através de uma atividade prática semelhante ao que era pretendido pelo experimento. Cada participante foi orientado a proceder conforme a descrição da tarefa e a ler quatro frases de ambientação (*We know English.; I hit table things.; You watch king shows.; I got televisions to fix.*).

Ao completar a tarefa, os sujeitos partiram à efetiva coleta de amostras e, a partir disso, foram orientados a não interromper a leitura das frases até que se chegasse ao final do experimento. A gravação das amostras obedeceu aos procedimentos detalhados na seção 3.1.3, “Gravações e coleta de dados do estudo piloto”. Cada participante teve seu *output* sonoro codificado e armazenado em arquivo de áudio para que se pudesse conduzir análises estatísticas com base na observação dos padrões de inteligibilidade.

A progressão do experimento foi marcada a cada 10 enunciados lidos, mas sem o uso de indicadores numéricos, para evitar a ansiedade dos participantes e manter a qualidade de produção sonora até o fim do experimento. A fim de manter um diálogo mais informal e tentar diminuir a tensão dos participantes durante a coleta, os indicadores de progressão

---

<sup>32</sup> O dispositivo usado para a coleta das amostras de fala desta pesquisa foi o mesmo usado à época de pilotagem dos experimentos.



obedeceram a um registro mais coloquial, conforme se demonstra no Quadro 09 que segue.

Quantidade de frases lidas	Indicador de progressão do experimento
10	Já se foram 10! Continue com o bom trabalho!
20	Estamos na vigésima! Continue assim!
30	Quase 1/3! Falta menos do que parece!
40	Já se foram mais 10! Continue com o bom trabalho!
50	Não desista! Estamos quase lá!
60	Respira fundo e vai! Já passamos de 2/3!
70	Falta pouco! Não desista!
80	Estamos na reta final! O último suspiro!
90	Agente firme! Continue com o bom trabalho!
100	Sinta o aroma! É a linha de chegada!
108	Acabou! Você venceu o desafio!

QUADRO 09 – Indicadores de progressão usados no experimento de coleta de amostras de fala.

Para a gravação do *corpus*, os participantes não levaram mais que 10 minutos cada um, conforme o Gráfico 02, e a média de tempo levada para a produção de cada frase foi de 4,9 segundos.

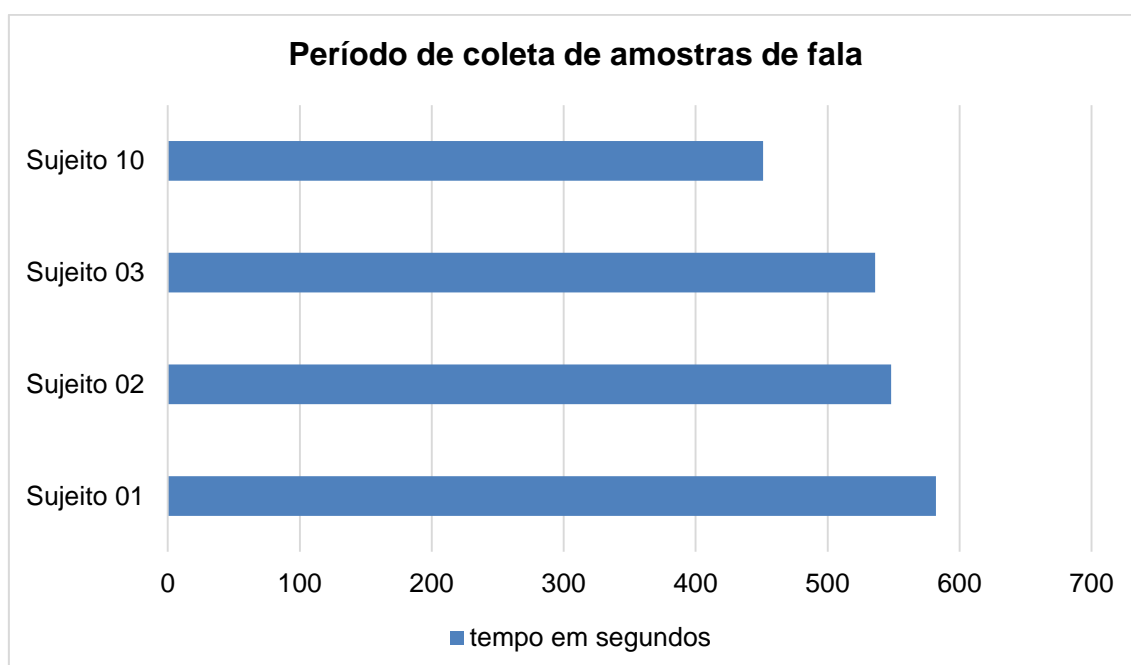


GRÁFICO 02 – Tempo (em segundos) demandado para a coleta das amostras de fala com o grupo de falantes.

A etapa seguinte à coleta de dados de produção dos participantes brasileiros consistiu em aplicar em arquivo de vídeo os áudios armazenados, para que pudessem ser importados para conta do YouTube no modo privado, protegendo, assim, a privacidade dos dados dos sujeitos. A transcrição dos arquivos seguiu os procedimentos de transcrição automática disponíveis no canal social: seleção do vídeo a ser transcrito, determinação do idioma do arquivo e inicialização da ferramenta. Optou-se pela transcrição em Inglês americano, conforme disponibilidade da ferramenta, uma vez que ambas as escolas em que os participantes estavam matriculados empregavam esta variante e os participantes alegaram, também, que identificam mais a sua fala com o inglês americano. Um breve tutorial de como proceder para transcrição de vídeos nessa plataforma foi disponibilizada no Apêndice G desta dissertação.

O YouTube oferece a possibilidade de exportação do arquivo de texto transcrito em formato compatível a softwares de leitura de texto (.srt), como o Bloco de Notas. A transcrição traz informações sobre a minutagem de cada frase (tempo de início e término do *input* sonoro em relação à transcrição), como se mostra na Figura 06.

FIGURA 06 – Relação entre minutagem e transcrição do *input* sonoro.

```

20
00:02:22,560 --> 00:02:33,209
to peruse your weights peanuts yesterday

21
00:02:27,360 --> 00:02:39,989
period we ship systems for you period

22
00:02:33,209 --> 00:02:45,180
I keep coffee here period i got cubes to

```

---

Para a análise prévia da inteligibilidade das amostras sonoras para a ferramenta de transcrição do YouTube, migrou-se as informações do arquivo .srt para uma planilha, em que traçou-se a relação entre o que foi transcrito e o

que foi enunciado pelos quatro participantes do grupo de falantes. A exposição e a análise desse cruzamento em função das diversas variáveis envolvidas (familiaridade lexical, anos de instrução e gênero, por exemplo) são descritas no Capítulo 4 desta dissertação.

O experimento de inteligibilidade foi desenvolvido com a intenção de verificar se a fala dos brasileiros que compuseram o grupo de falantes era inteligível ao ouvido humano. Essa intenção se justifica porque se viu necessária a triangulação dos dados coletados. Este experimento, apesar de delineado, não foi aplicado à época de pilotagem, pois não se pôde contar com a contribuição de informantes que atendessem às especificidades biográficas estipuladas para a pilotagem.

A variedade americana do idioma estrangeiro usado como base para a coleta foi determinada como parâmetro para a ferramenta de reconhecimento usada posteriormente ao experimento de coleta de amostras de fala (YouTube). Por isso a necessidade de se ter um grupo controle de falantes nativos de inglês americano, para manter a compatibilidade das variedades do inglês empregado pelo aplicativo e pela instituição de ensino em que os voluntários brasileiros estavam matriculados. Entretanto, no contexto de Língua Franca discutido no Capítulo 2 desta dissertação, optou-se por investigar a inteligibilidade das amostras sonoras do grupo de falantes, também, para ouvintes de outras línguas maternas com trajetórias linguísticas diferentes, como foi evidenciado no Quadro 07, a saber, 2 ouvintes brasileiros e 2 americanos.

Para que não houvesse divergência entre o *input* que o aplicativo recebeu e o que os participantes ouvintes teriam contato, as amostras de fala coletadas do grupo de falantes foram fragmentadas e usadas para a confecção do experimento de inteligibilidade. Sabe-se que a reprodução idêntica de um segmento é inviável de se esperar, por isso fez-se essa escolha metodológica, que ofereceria garantias de que ambos os grupos de ouvintes (aplicativo e humano) seriam expostos a exemplares de natureza idêntica.

A tarefa de inteligibilidade seguiria a proposta assumida em Becker (2013), em que os ouvintes seriam orientados a transcrever as frases a que eram expostos. O experimento foi elaborado usando a ferramenta online *Google Forms*. Com a possibilidade de inserção de vídeo, a composição do

teste foi relativamente simples e funcionou sem fatores que pudessem interromper a condução do experimento.

A seleção das frases empregadas no experimento obedeceu a alguns critérios. Foram descartadas as primeiras três ocorrências de cada palavra-alvo (das seis produzidas por cada participante) e selecionadas a quarta ocorrência de cada frase para cada participante. Para não exaurir os potenciais ouvintes, optou-se por restringir o número de frases a serem transcritas, de modo que cada ouvinte não teria que transcrever mais que 21 sentenças. Tais estímulos foram recortados das amostras coletadas durante o experimento de interação com o aplicativo. Haja vista que inicialmente sete participantes submeteram-se à coleta de produção, determinou-se a escolha de uma amostra para cada segmento consonantal (/p, t, k/) por participante. Por isso, foram selecionadas três frases produzidas por cada participante brasileiro de modo que todas as palavras-alvo fossem contempladas e transcritas pelos ouvintes no experimento de inteligibilidade. O Quadro 10 mostra quais sentenças de cada participante do grupo de falantes foram expostas ao grupo de ouvintes.

Sujeito	Amostra	Palavra-alvo	Frase
01	/k/	<i>keepers</i>	<i>We help keepers downstairs.</i>
	/p/	<i>pizza</i>	<i>We eat pizza for lunch.</i>
	/t/	<i>teaspoons</i>	<i>We ship teaspoons for you.</i>
02	/t/	<i>teaser</i>	<i>We talk teaser sentences.</i>
	/p/	<i>peanuts</i>	<i>We ate peanuts yesterday.</i>
	/k/	<i>keep</i>	<i>We keep coffee here.</i>
03	/p/	<i>people</i>	<i>We hit people downstairs.</i>
	/k/	<i>keels</i>	<i>We got keels to fix.</i>
	/t/	<i>teens</i>	<i>We help teens at school.</i>
04	/p/	<i>peaks</i>	<i>We walk peaks together.</i>
	/t/	<i>teachers</i>	<i>We help teachers at school.</i>
	/k/	<i>keyboards</i>	<i>We got keyboards to fix.</i>
06	/p/	<i>peace</i>	<i>We got peace like a river.</i>
	/k/	<i>keys</i>	<i>We got keys downstairs.</i>
	/t/	<i>team</i>	<i>We stop team fights.</i>
08	/p/	<i>peach</i>	<i>We ate peach yesterday.</i>
	/k/	<i>kiwi</i>	<i>We eat kiwi for lunch.</i>
	/t/	<i>teeth</i>	<i>We help teeth problems.</i>
10	/p/	<i>pizza</i>	<i>We eat pizza for lunch.</i>

	<b>/k/</b>	<b><i>keepers</i></b>	<b><i>We help keepers downstairs.</i></b>
	<b>/t/</b>	<b><i>teaspoons</i></b>	<b><i>We ship teaspoons for you.</i></b>

QUADRO 10 – Seleção de frases empregadas no teste de inteligibilidade.

Cabe esclarecer que o experimento de inteligibilidade com o grupo de ouvintes foi feito antes de se submeter os dados de produção de todos os 7 participantes à transcrição automática. Como a ferramenta YouTube transcreveu os dados de apenas 4 dos 7 participantes, o experimento de inteligibilidade com os ouvintes teve uma redução de número de dados, uma vez que, para fins comparativos, somente os dados dos 4 participantes que foram efetivamente transcritos pelo YouTube foram analisados no experimento de inteligibilidade dos ouvintes. Assim, para o experimento de inteligibilidade, foram analisados somente as frases (contendo os segmentos/vocábulo-alvo) que foram selecionadas para os 4 participantes que também foram transcritos pelo YouTube, a saber participantes 01, 02, 03 e 10 que estão em negrito no Quadro 10.

Desta forma, foram analisadas 4 amostras de cada um dos segmentos-alvo (/p, t, k/), para cada um dos quatro participantes ouvintes, totalizando 48 *tokens*, 16 para cada segmento e distribuídos nas palavras-chave *pizza* (2 amostras), *peanuts*, *people*, *teaspoons* (2 amostras), *teaser*, *teens*, *keep*, *keels*, *keepers* (2 amostras).

Da mesma maneira que o aplicativo, os participantes do grupo de ouvintes tiveram a chance de ouvir apenas uma vez ao *input* sonoro e foram orientados a transcrever conforme o que ouviram. Ao final, com o formulário encerrado, suas respostas foram armazenadas em planilha, conforme o padrão disponibilizado pela ferramenta de criação de formulários.

Ao grupo de ouvintes também foi aplicado o questionário de familiaridade. A opção de aplicação dos experimentos a esse grupo na ordem em que foi feito se justifica pela intenção de não influenciar o julgamento de inteligibilidade dos participantes do grupo. A familiaridade das palavras-alvo para o grupo de ouvintes se demonstra no Quadro 11. No quadro, “sim” indica que a palavra é familiar ao participante e “não” indica que o vocábulo não o é ao participante.

<b>Palavra-alvo</b>	<b>01.1</b>	<b>02.1</b>	<b>03.1</b>	<b>04.1</b>
---------------------	-------------	-------------	-------------	-------------

<i>peach</i>	sim	sim	sim	sim
<i>peace</i>	sim	sim	sim	sim
<i>peak</i>	sim	<b>não</b>	sim	sim
<i>pizza</i>	sim	sim	sim	sim
<i>peanut</i>	sim	sim	sim	sim
<i>people</i>	sim	sim	sim	sim
<i>key</i>	sim	sim	sim	sim
<i>keep</i>	sim	sim	sim	sim
<i>keel</i>	sim	<b>não</b>	sim	sim
<i>kiwi</i>	sim	sim	sim	sim
<i>keyboard</i>	sim	sim	sim	sim
<i>keeper</i>	sim	sim	sim	sim
<i>teeth</i>	sim	sim	sim	sim
<i>teen</i>	sim	sim	sim	sim
<i>team</i>	sim	sim	sim	sim
<i>teacher</i>	sim	sim	sim	sim
<i>teaspoon</i>	sim	sim	sim	sim
<i>teaser</i>	sim	<b>não</b>	sim	sim

QUADRO 11 – Familiaridade das palavras-alvo para o grupo de ouvintes.

Observando-se o Quadro 11, pode-se dizer que todas as palavras-alvo são familiares para os ouvintes americanos (03.1 e 04.1). Para os ouvintes brasileiros, o sujeito 02.1 relatou que as palavras *peak*, *keel* e *teaser* não lhe são familiares.

O impacto que os dados coletados apresentaram para a inteligibilidade dos segmentos oclusivos surdos /p/, /t/, e /k/ e dos vocábulos-alvo em que estão inseridos para a ferramenta de transcrição do YouTube e para o grupo de ouvintes será discutido no capítulo que segue em análise estatística e descritiva.

## 4 APRESENTAÇÃO E ANÁLISE DOS DADOS

Nesta seção, os dados coletados serão expostos e será apresentada uma análise de inteligibilidade de cada segmento sonoro em evidência (/p, t, k/) assim como as palavras-alvo em que estão inseridos. Para efeitos de organização, primeiro serão apresentadas as análises estatísticas que demonstram a inteligibilidade de cada segmento e lexema para a ferramenta de transcrição utilizada, recorrendo-se apenas sobre os casos em que os resultados dos testes estatísticos foram significativos para inteligibilidade ( $p < 0,05$ ).

A análise quantitativa foi conduzida por meio do *software Statistical Package for Social Studies®* (SPSS, versão 17.0) e foram rodados testes não paramétricos uma vez que os dados não estavam distribuídos normalmente. Em seguida, será conduzida a apresentação dos resultados relativos ao experimento de inteligibilidade do grupo de ouvintes, bem como das ocorrências em que se percebeu um padrão de reconhecimento (ou não) tanto da ferramenta de transcrição quanto do grupo de ouvintes.

### 4.1A INTELIGIBILIDADE PARA A FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE

Análises quantitativas são, em geral, apresentadas quando há um número expressivo de dados que possam ser postos em contraste. Em função das dificuldades descritas no capítulo anterior, o volume de dados pretendidos não pode ser alcançado. Apesar disso, a metodologia aplicada permitiu que uma quantidade suficiente de dados pudesse ser coletada e, por conseguinte, analisada estatisticamente. O Quadro 12 ilustra a distribuição dos dados coletados.

Sujeitos	Segmentos	Palavras-alvo	Total de repetições por palavra-alvo	Total de repetições por segmento	Total de amostras por sujeito	Total geral
4	/p/	peace	6	36	108	432

		<i>peaks</i> <i>pizza</i> <i>peanuts</i> <i>peach</i> <i>people</i>				
	/t/	<i>teacher</i> <i>team</i> <i>teaser</i> <i>teens</i> <i>teaspoons</i> <i>teeth</i>	6	36		
	/k/	<i>keels</i> <i>keepers</i> <i>keep</i> <i>keyboards</i> <i>keys</i> <i>kiwi</i>	6	36		

QUADRO 12 – Distribuição quantitativa dos dados coletados.

Conforme apresentado no Capítulo 3 dessa dissertação, cada sujeito do grupo de aprendizes brasileiros produziu seis amostras de cada palavra-alvo. Como somente a produção de quatro sujeitos do grupo de falantes foram efetivamente transcritas pela ferramenta do YouTube, totaliza-se 432 das 1512 amostras pretendidas para análise, distribuídas igualmente entre cada segmento oclusivo (144 para /p/, 144 para /t/ e 144 para /k/, sendo, para cada segmento, 6 palavras-alvo x 6 repetições x 4 participantes, totalizando 144).

Os valores médios de inteligibilidade para a ferramenta de transcrição automática do YouTube foram calculados em função da quantidade de vezes em que foram reconhecidos por tal ferramenta. Isto é, para cada sujeito, foram contabilizadas as ocorrências de reconhecimento do segmento oclusivo surdo e as ocorrências em que os segmentos não foram inteligíveis. Ainda que o foco da investigação recaia sobre os segmentos em si, as palavras-alvo também foram analisadas para que os possíveis níveis de interferência das oclusivas na inteligibilidade lexical pudessem ser apontados.

O Quadro 13 mostra as médias e o desvio padrão de inteligibilidade dos segmentos oclusivos para a ferramenta de transcrição do YouTube, assim



como a variação de distribuição entre os participantes do grupo de aprendizes brasileiros.

<b>Segmento oclusivo</b>	<b>Média de inteligibilidade (acertos)</b>	<b>Reconhecimento mínimo</b>	<b>Reconhecimento máximo</b>
[p]	18,75	9	27
[t]	11,50	7	19
[k]	27,25	17	33

QUADRO 13 – Média de inteligibilidade e distribuição dos segmentos oclusivos surdos produzidos pelo grupo de aprendizes brasileiros para a ferramenta de transcrição do YouTube.

O Quadro 13 tem por base a produção de 36 *tokens* por participante e evidencia uma possível ordem de inteligibilidade entre os segmentos. O teste de Friedman indicou haver diferenças estatisticamente significativas entre a inteligibilidade das três consoantes ( $p < 0,05$ ), em que [k] é mais inteligível que [p], que é mais inteligível que [t]. No entanto, os Testes de Wilcoxon, com correção de Bonferroni, não sugeriram diferenças estatisticamente significativas entre nenhum dos pares [p] e [t] ( $Z = -1,826$ ;  $p > 0,05$ ), [p] e [k] ( $Z = -1,841$ ;  $p > 0,05$ ) e [t] e [k] ( $Z = -1,826$ ;  $p > 0,05$ ). Apesar dos resultados entre os pares não serem significativos, observamos uma tendência a serem marginalmente significativos, tendência que poderia ser confirmada caso o número de dados analisados fosse ampliado, por exemplo.

Em relação aos vocábulos-alvo iniciados em /p/, /t/ e /k/, percebeu-se relação semelhante de ranqueamento em comparação à inteligibilidade dos segmentos em si para a ferramenta de transcrição do YouTube, ou seja, os vocábulos iniciados por [k] são mais inteligíveis que os iniciados por [p], que são mais inteligíveis que os vocábulos iniciados por [t], como ilustra o Quadro 14.

<b>Segmento oclusivo em onset</b>	<b>Média de inteligibilidade (acertos)</b>	<b>Reconhecimento mínimo</b>	<b>Reconhecimento máximo</b>
[p]	14,50	7	21
[t]	10,50	5	18
[k]	16,00	7	21

QUADRO 14 – Média de inteligibilidade da ferramenta de transcrição automática do YouTube e distribuição dos vocábulos iniciados por segmentos oclusivos surdos produzidos pelo grupo de aprendizes brasileiros.

O teste de Friedman indicou haver diferenças estatisticamente significativas entre a inteligibilidade das palavras-alvo iniciadas pelas três consoantes ( $p < 0,05$ ). Os testes de Wilcoxon conduzidos, com correção de Bonferroni, comparando as três variáveis, agrupadas duas a duas, não sugeriram diferenças estatisticamente significativas, seguindo o mesmo que ocorreu para a análise dos segmentos-alvo.

Comparando os resultados obtidos sobre a inteligibilidade dos segmentos oclusivos surdos e dos lexemas iniciados por esses segmentos para a ferramenta de transcrição do YouTube, percebe-se que os segmentos são mais inteligíveis que os vocábulos, como indica o Quadro 15, que retoma as médias de inteligibilidade de [p, t, k] quando só fonema e quando inseridos em palavra-alvo.

	Fonema	Palavra-alvo
[k]	27,25	16,00
[p]	18,75	14,50
[t]	11,50	10,50

QUADRO 15 – Médias de inteligibilidade dos segmentos oclusivos surdos e dos vocábulos iniciados por estes para a ferramenta de transcrição automática do YouTube.

Em ambos os cenários parece que há uma tendência do mais inteligível para o menos inteligível ( $[k > p > t]$ ). Entretanto, a diferença nos valores médios de inteligibilidade na comparação fonema x léxico mostra que os segmentos consonantais em si são mais inteligíveis que as palavras em que são inseridos. Isso pode se explicar pelo fato de que no vocábulo outros segmentos são considerados para a inteligibilidade e que a realização destes outros segmentos da palavra (duração da vogal, aspiração da consoante oclusiva, etc.) podem ter interferido na inteligibilidade da palavra como um todo e não somente os segmentos-alvo.

O detalhamento desses dados será apresentado na sequência, em que os segmentos e palavras-alvo serão analisados e discutidos separadamente conforme o ponto de articulação das consoantes investigadas.

#### 4.1.1 A oclusiva surda bilabial

Na iniciativa de medir a inteligibilidade de cada segmento oclusivo (isolado ou inserido em palavra-alvo), para cada transcrição realizada pela ferramenta automática do YouTube, contabilizou-se as vezes em que o mecanismo reconheceu o segmento consonantal (neste caso o /p/) e as vezes em que este não foi inteligível. Da mesma maneira, quantificou-se as ocorrências em que as palavras-alvo iniciadas pela consoante bilabial surda foram inteligíveis ou não. Cada palavra-alvo iniciada por [p] foi transcrita pelo YouTube 24 vezes (palavra-alvo x 6 repetições x 4 participantes, totalizando 24).

O contraste desses valores (em que [p] foi inteligível e em que [p] não o foi) foi determinado por meio de testes de Wilcoxon, tendo como valor de referência de significância estatística  $p < 0,05$ . Os resultados foram simplificados e inseridos no Quadro 16 para melhor ilustrar as relações de inteligibilidade, apresentando-se, também, as médias de inteligibilidade (indicadas conforme o número máximo de 6 amostras para cada sujeito do grupo de aprendizes brasileiros).

<b>Segmento-alvo</b>	<b>Número de vezes inteligível</b>	<b>Número de vezes não inteligível</b>	<b>Média de inteligibilidade (acertos)</b>	<b>Valor de p</b>
[p] em <i>peace</i>	13	11	3,25	0,854 <sup>a</sup>
[p] em <i>peach</i>	2	22	0,50	0,063 <sup>b</sup>
[p] em <i>peaks</i>	11	13	2,75	0,854 <sup>b</sup>
[p] em <i>peanuts</i>	10	14	2,50	0,581 <sup>b</sup>
[p] em <i>people</i>	22	2	5,50	0,066 <sup>a</sup>
[p] em <i>pizza</i>	17	7	4,25	0,144 <sup>a</sup>
TOTAL	75	69	18,75	0,715 <sup>a</sup>

QUADRO 16 – Média de inteligibilidade do segmento oclusivo bilabial surdo /p/ para a ferramenta de reconhecimento do YouTube.

O segmento [p] foi avaliado conforme a inserção em palavra-alvo, ou seja, cada linha do Quadro 16 representa, estatisticamente, o quão significativa foi a inteligibilidade do segmento para a ferramenta de transcrição automática

do YouTube. Os valores de  $p$  indicam que o segmento oclusivo bilabial surdo não é estatisticamente significativo para a inteligibilidade, ou seja, a análise estatística não revela uma tendência de reconhecimento para nenhuma das palavras iniciadas pelo segmento-alvo. Entretanto, conforme apontando previamente, o baixo volume de dados analisados pode ter corroborado para esse resultado, uma vez os resultados de inteligibilidade para as palavras *peach* e *people* podem ser considerados marginalmente significativos uma vez que o valor de  $p$  está mais próximo de 0,05.

Comparando os valores de significância estatística do fone [p] em *peach* e em *people*, pode-se perceber um movimento que os aproxima do valor de referência que os tornaria significativos à inteligibilidade ( $p < 0,05$ ). Em contraste, [p] em *peach* atingiu uma média de inteligibilidade baixa (0,50 para cada 6 ocorrências), tendendo a não ser inteligível a ferramenta YouTube, enquanto [p] em *people* tende ser inteligível (5,50 para cada 6 ocorrências) a mesma ferramenta de reconhecimento.

Na avaliação dos efeitos de [p] para a inteligibilidade das palavras-alvo enquanto lexemas (isto é, se os vocábulos iniciados em /p/ foram reconhecidos pela ferramenta de transcrição do YouTube) percebeu-se uma tendência que se assemelha à observada na análise do segmento isolado, ilustrada no Quadro 17.

Palavra-alvo	Número de vezes inteligível	Número de vezes não inteligível	Média de inteligibilidade (acertos)	Valor de p
<i>peace</i>	13	11	3,25	0,854 <sup>a</sup>
<i>peach</i>	1	23	0,25	0,059 <sup>b</sup>
<i>peaks</i>	2	22	0,50	0,059 <sup>b</sup>
<i>peanuts</i>	4	20	1,00	0,102 <sup>b</sup>
<i>people</i>	22	2	5,50	0,066 <sup>a</sup>
<i>pizza</i>	16	8	4,00	0,273 <sup>a</sup>
TOTAL	58	86	14,50	0,465 <sup>a</sup>

QUADRO 17 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo bilabial surdo /p/ para a ferramenta de reconhecimento do YouTube.

Os valores indicados para  $p$  indicam que a inteligibilidade de *peach* e *peaks* é marginalmente significativa. Isso significa que, em ambos os casos, as

palavras-alvo tiveram baixa taxa de reconhecimento (média de 0,25 para a primeira e 0,50 para a segunda). Em aprofundamento, *peach* se mostrou inteligível em, no máximo, uma amostra dos participantes do grupo de aprendizes brasileiros, e *peaks* em duas para a ferramenta de transcrição do YouTube.

Ao observar os dados de transcrição do YouTube, percebe-se uma tendência ao reconhecimento da palavra *peach* como *speech*. Semanticamente, espera-se que a ferramenta de transcrição busque por um vocábulo-alvo que seja compatível à sequência frasal. Em alguns casos, o verbo *eat* (que precedia a palavra-alvo) foi reconhecido pela ferramenta, mas o substantivo que o segue não. O Quadro 18 traz a transcrição dessas frases em que a palavra-alvo *peach* está inserida, mas foi reconhecida como *speech*.

Sujeito	Transcrição do YouTube
01	I ate speech yesterday period
	I eat speech yesterday period
	with 8 speech yesterday period
	you ate speech yesterday period
	you ate speech yesterday two years
	we make speech yesterday to
02	you a speech yesterday cute
	we ate speech yesterday period
	I ate speech yesterday to you
03	you a speech yesterday period
	we in speech yesterday period
10	you a speech yesterday spirit

QUADRO 18 – Transcrição do YouTube para as frases em que houve a inserção da fricativa [s] anteriormente ao segmento [p] na palavra-alvo *peach*.

A tendência observada pode ter sido resultante de uma palatalização da consoante em *coda* do verbo *eat* (/t/). Como não foram conduzidas análises acústicas para verificação deste fenômeno (somente de outiva), soma-se a isso a possível interferência que o ruído da gravação tenha causado, ocasionando, possivelmente a inserção de uma fricativa (/s/) no início da palavra-alvo pela ferramenta YouTube.

Em análise qualitativa direcionada ao vocábulo-alvo *peaks* ( $p < 0,05$ ), percebeu-se a recorrência do vozeamento da consoante posterior à vogal

controlada (/i:/). Pensa-se que esse vozeamento pode ser atribuído ao encurtamento da vogal (/ɪ/ ao invés de /i:/) e à aspiração da consoante surda (neste caso, /k/). A distância dos pontos de articulação da vogal e da consoante posterior a ela (a vogal é anterior e a consoante é velar) pode ter causado o encurtamento da vogal e determinado esse padrão de transcrição pela ferramenta disponível no YouTube. O Quadro 19 mostra as ocorrências desse fenômeno, que se repetiu em 11 das 24 amostras com a palavra-alvo *peaks*.

Sujeito	Transcrição do YouTube
01	we walk pigs together period
	you walk pigs together to use
	i walk pigs together period
	we all win war pigs together to you
	you walk pigs together to use
02	I walk pigs together do
	we walk pigs together cute
	you are big to get a student
	I won't dig together period
03	we walk pigs together period
10	you will dig together period

QUADRO 19 – Transcrição do YouTube para as frases em que houve o vozeamento da consoante /k/ na palavra-alvo *peaks*.

Inversamente, *people* aparenta ser mais inteligível à ferramenta de transcrição, indicando  $p = 0,066$ , o que se aproxima marginalmente do valor de referência. O Gráfico 03 traz a distribuição de reconhecimento dos vocábulos *peach*, *peaks* e *people* para cada sujeito do grupo de falantes.

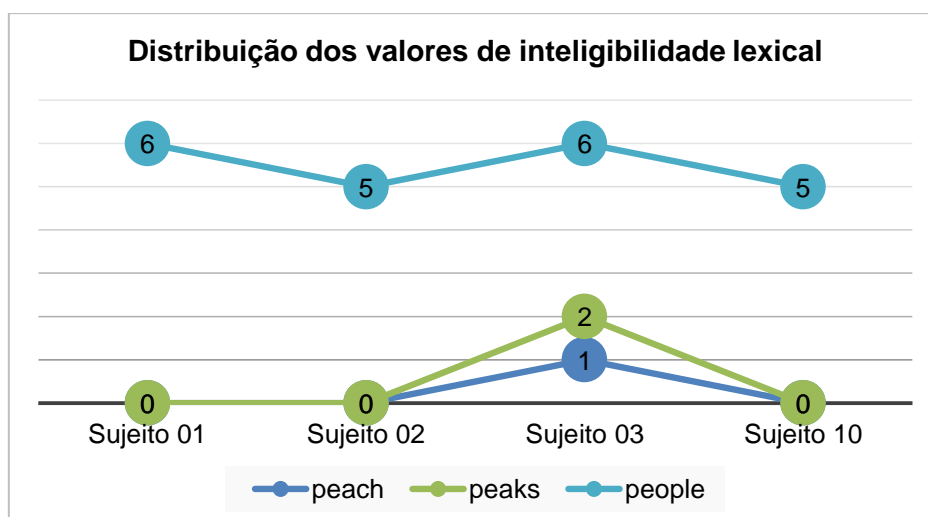


GRÁFICO 03 – Distribuição dos valores de inteligibilidade das palavras-alvo *peach*, *peaks* e *people* para a ferramenta de transcrição do YouTube.

Essa diferença de inteligibilidade, apesar de não ser classificada como estatisticamente significativa, pode ter sido causada, também, pela frequência de uso dos vocábulos. Na lista das 5000 palavras mais frequentes disponibilizada no *corpus* COCA<sup>33</sup>, *people* aparece na 62ª posição, seguido por *peak*, ocupando a posição 2385. *Peach* não aparece na lista disponibilizada. Não se acredita que a familiaridade dos vocábulos *peach* e *peaks* tenha influenciado o reconhecimento destes pela ferramenta de transcrição automática do YouTube. No formulário de familiaridade aplicado aos falantes, apenas um deles reportou não reconhecer nem saber o significado da palavra-alvo *peaks*, enquanto os vocábulos *peach* e *people* foram reportados como familiares a todos os sujeitos do grupo de falantes.

#### 4.1.2 A oclusiva surda alveolar

Para a análise estatística da consoante [t] em posição de ataque silábico, quantificou-se as ocorrências de reconhecimento da ferramenta de transcrição automática do YouTube tanto para o segmento isolado quanto para a palavra-alvo. Para cada segmento e para cada palavra-alvo, obteve-se seis amostras de cada aprendiz brasileiro, totalizando 24 (1 palavra x 6 repetições x 4 participantes). O Gráfico 04 mostra a distribuição de inteligibilidade do segmento oclusivo alveolar.

<sup>33</sup> Corpus of Contemporary American English (COCA). Disponível em: <www.wordfrequency.info> Acesso em: 15 Ago. 2017.

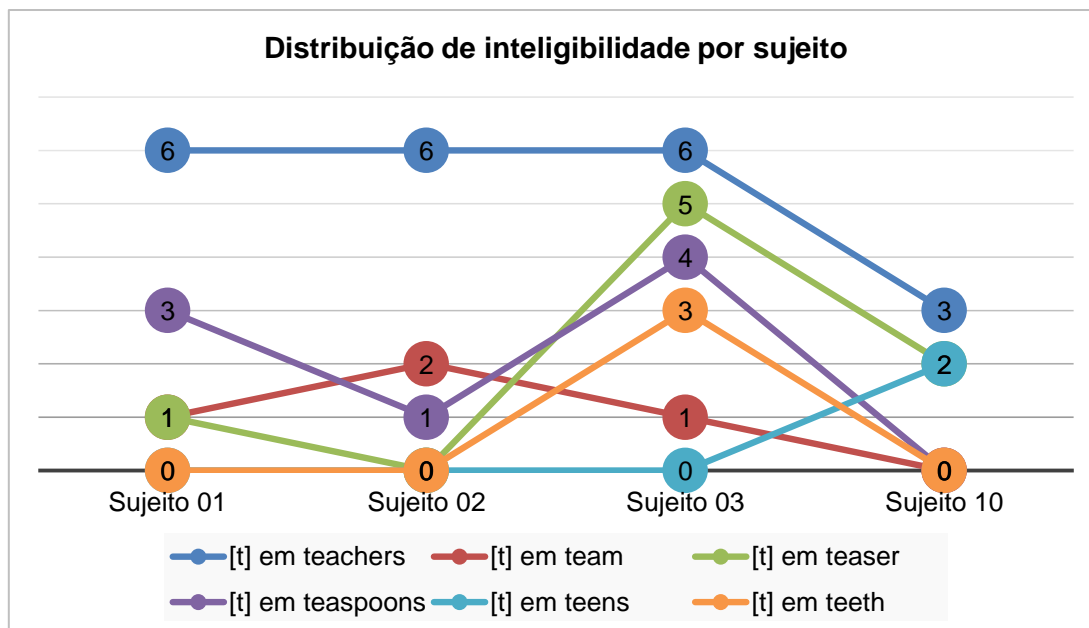


GRÁFICO 04 – Distribuição de inteligibilidade da consoante oclusiva alveolar por aprendiz brasileiro para a ferramenta de transcrição do YouTube.

O movimento do gráfico aponta para a baixa inteligibilidade de [t] em *team*, *teens* e *teeth*. Apesar das diferenças apontadas, testes de Wilcoxon indicaram que há diferença estatística marginalmente significativa apenas para [t] em *teens* ( $p = 0,059$ ). O segmento [t] em *team* também se aproximou do valor de significância tido como referência ( $p < 0,05$ ), mas o volume de dados coletados não permite que se afirme ter havido correlação entre os fenômenos que causaram a baixa inteligibilidade do segmento consonantal inicial em *teens* e em *team*. O Quadro 20 descreve a média de inteligibilidade da consoante [t] em cada palavra-alvo para a ferramenta de transcrição do YouTube. Cada palavra alvo iniciada por [t] foi transcrita pelo YouTube 24 vezes (palavra-alvo x 6 repetições x 4 participantes).

Segmento-alvo	Número de vezes inteligível	Número de vezes não inteligível	Média de inteligibilidade (acertos)	Valor de p
[t] em <i>teacher</i>	21	3	5,25	0,083 <sup>a</sup>
[t] em <i>team</i>	4	20	1,00	0,066 <sup>b</sup>
[t] em <i>teaser</i>	8	16	2,00	0,357 <sup>b</sup>
[t] em <i>teaspoons</i>	8	16	2,00	0,285 <sup>b</sup>
[t] em <i>teens</i>	2	22	0,50	0,059 <sup>b</sup>
[t] em <i>teeth</i>	3	21	0,75	0,131 <sup>b</sup>



TOTAL	46	98	11,50	0,715 <sup>a</sup>
-------	----	----	-------	--------------------

QUADRO 20 – Média de inteligibilidade do segmento oclusivo alveolar surdo /t/ para a ferramenta de reconhecimento do YouTube.

O Quadro 21 detalha as médias de inteligibilidade das palavras-alvo iniciadas em /t/ e demonstra o valor de significância dos dados coletados.

Palavra-alvo	Número de vezes inteligível	Número de vezes não inteligível	Média de inteligibilidade (acertos)	Valor de p
<i>teacher</i>	21	3	5,25	0,083 <sup>a</sup>
<i>team</i>	3	21	0,75	0,066 <sup>b</sup>
<i>teaser</i>	8	16	2,00	0,357 <sup>b</sup>
<i>teaspoons</i>	7	17	1,75	0,180 <sup>b</sup>
<i>teens</i>	0	24	0,00	0,046 <sup>b</sup>
<i>teeth</i>	3	21	0,75	0,131 <sup>b</sup>
TOTAL	42	102	10,50	0,144 <sup>a</sup>

QUADRO 21 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo alveolar surdo /t/ para a ferramenta de reconhecimento do YouTube.

Na avaliação dos efeitos de [t] para a inteligibilidade das palavras-alvo, percebeu-se uma tendência que se assemelha à observada na análise do segmento isolado. Testes de Wilcoxon indicam que a falta de inteligibilidade da palavra *teens* é estatisticamente significativa ( $p < 0,05$ ), tendo sido reconhecida em nenhuma das amostras expostas à ferramenta de transcrição do YouTube.

Buscando entender os recursos utilizados pelo mecanismo de transcrição da ferramenta do YouTube, percebeu-se a repetição do deslocamento do segmento-alvo (/t/) da posição de *onset* do vocábulo-alvo para a posição de *coda* do verbo o precedia. Esse movimento de migração pode ter ocorrido porque o *software* opera com probabilidades, comparando e contrastando o seu banco de dados (inventário fonético e sintático) com o *input* sonoro oferecido. O Quadro 22 evidencia a ocorrência desse fenômeno, evidenciado em 11 das 24 amostras de fala transcritas.

Sujeito	Transcrição do YouTube
01	we helped in the school period

<sup>a</sup> Baseado nos valores positivos (em que o fonema foi reconhecido pela ferramenta).

<sup>b</sup> Baseado nos valores negativos (em que o fonema não foi reconhecido pela ferramenta).

	I helped in the school period
	we helped in the school period
	you helped in the school period
02	I helped in the school give it
	we helped in the school do
	I helped King the school period
03	we helped things that whole period
	I helped in that call period
	you helped in the scope period
10	we helped in a school period

QUADRO 22 – Transcrição do YouTube para as frases em que houve o deslocamento da consoante /t/ para a posição de *coda* do vocábulo anterior à palavra-alvo *teens*.

O deslocamento da consoante em *onset* para *coda* teve impacto, também, no reconhecimento da ferramenta de transcrição para o segmento /t/ isolado quando inserido na palavra-alvo *teens*. Sabe-se que a qualidade da consoante em ataque silábico e em *coda* diferem, mas pode-se dizer que, apesar dessas diferenças, o segmento foi reconhecido pelo *software* de transcrição automática, mesmo que não no contexto em que se esperava.

Apesar de o resultado do teste Wilcoxon para a inteligibilidade da palavra-alvo *team* não ter sido considerado estatisticamente significativo ( $p = 0,66$ ), observou-se o mesmo padrão de reconhecimento evidenciado para *teens*, em que houve o deslocamento do segmento-alvo para a posição de *coda* do vocábulo precedente. O Quadro 23 traz a relação das ocorrências desse fenômeno para as frases transcritas em que a palavra-alvo *team* foi empregada.

Sujeito	Transcrição do YouTube
01	I stopped in fight period
	I stopped a fight period
	you stopped in fight period
	you stopped in fight career
02	you stopped in height cute
	we stopped in fact stood
	I stopped inside spirit
	I stopped in height period
03	you stopped teamfight period
	I stopped in fight period

	we stopped in flight period
	is stopped in flight period

QUADRO 23 – Transcrição do YouTube para as frases em que houve o deslocamento da consoante /t/ para a posição de *coda* do vocábulo anterior à palavra-alvo *team*.

Em contraste, das palavras-alvo analisadas iniciadas em /t/ a que parece ser mais inteligível ao mecanismo de transcrição é o vocábulo *teacher*. A alta média de inteligibilidade (que ficou evidenciada, também, na média de inteligibilidade de [t] em *teacher*) do lexema pode ser atribuída a sua frequência de uso. Nenhum dos participantes do grupo de aprendizes brasileiros reportou não reconhecer ou saber o que significa o vocábulo (o que foi entendido como familiar a todos) e o *corpus* COCA traz o item lexical ranqueado na 372ª posição.

#### 4.1.3 A oclusiva surda velar

Das consoantes avaliadas, a que mostrou valores mais altos de inteligibilidade foi a oclusiva velar surda /k/. É sabido que quanto mais distante do ponto de articulação da consoante estiver a vogal que a sucede, maior é o valor de VOT (SCHWARTZHAUPT; ALVES, 2014). O ranqueamento apresentado anteriormente pode ter sido influenciado por esse fator. Ainda, dentre os vocábulos-alvo selecionados para a condução dos experimentos havia um que era pouco familiar ao grupo de falantes (*keels*) e outro que não atendia aos parâmetros determinados (substantivo iniciado em consoante oclusiva surda seguida de vogal anterior alta /i:/).

Essas particularidades serão desmembradas a partir da análise qualitativa dos dados. Como ponto de partida, optou-se pela apresentação dos dados brutos e da análise quantitativa para que se pudessem identificar os casos em que a inteligibilidade (ou falta dela) do segmento e dos lexemas iniciados pelo segmento aparenta ser estatisticamente significativa. A distribuição dos dados é ilustrada no Quadro 24. Cada palavra-alvo iniciada por [k] foi transcrita pelo YouTube 24 vezes (palavra-alvo x 6 repetições x 4 participantes, totalizando 24).

Segmento-alvo	Número de	Número de	Média de	Valor de
---------------	-----------	-----------	----------	----------

	<b>vezes Inteligível</b>	<b>vezes não inteligível</b>	<b>inteligibilidade (acertos)</b>	<b>p</b>
[k] em <i>keels</i>	13	11	3,25	0,655 <sup>b</sup>
[k] em <i>keep</i>	24	0	6,00	0,059 <sup>a</sup>
[k] em <i>keepers</i>	16	8	4,00	0,357 <sup>a</sup>
[k] em <i>keyboards</i>	20	4	5,00	0,102 <sup>a</sup>
[k] em <i>keys</i>	20	4	5,00	0,063 <sup>a</sup>
[k] em <i>kiwi</i>	15	9	4,00	0,285 <sup>a</sup>
TOTAL	108	36	27,25	0,144 <sup>a</sup>

QUADRO 24 – Média de inteligibilidade do segmento oclusivo velar surdo /k/ para a ferramenta de reconhecimento do YouTube.

Em comparação com as demais consoantes oclusivas surdas (/p/ e /t/), o segmento mostra ser mais inteligível à ferramenta de transcrição do YouTube. O volume de dados não permite dizer que a alta inteligibilidade do segmento seja estatisticamente significativa (em que  $p < 0,05$ ), mas evidencia que o segmento é reconhecido em  $\frac{3}{4}$  das ocorrências (75%).

É importante destacar que a inteligibilidade de [k] em *keep* mostra ser marginalmente significativa, aproximando-se do valor de referência. A alta inteligibilidade do segmento nesse contexto pode ter se dado em função da posição sintática que ocupou na frase. Ao contrário dos demais vocábulos, *keep* foi empregado como verbo transitivo direto e isso pode ter facilitado o trabalho de reconhecimento pela ferramenta de transcrição. Vale lembrar que esse tipo de mecanismo informático trabalha com probabilidades, tanto no nível lexical quanto no sintático, e que a alta frequência de uso do verbo *keep* (posição 156 conforme o COCA) pode ter sido determinante na escolha lexical para a ferramenta de transcrição. Da mesma maneira, a inteligibilidade de [k] em *keys* foi marginalmente significativa, tendendo a ser inteligível. O Quadro 24 mostra que o segmento foi reconhecido em 20 das 24 ocorrências.

Com o propósito de poder investigar a interferência do segmento oclusivo velar para a inteligibilidade dos lexemas iniciados por este, conduziu-se testes de Wilcoxon, indicando as vezes em que os vocábulos foram inteligíveis e as em que não os foram, conforme é apresentado no Quadro 25.

<sup>a</sup> Baseado nos valores positivos (em que o fonema foi reconhecido pela ferramenta).

<sup>b</sup> Baseado nos valores negativos (em que o fonema não foi reconhecido pela ferramenta).

O Quadro também traz a média de inteligibilidade (com valor de referência para cada 6 ocorrências) e os valores de p para cada vocábulo.

Palavra-alvo	Número de vezes inteligível	Número de vezes não inteligível	Média de inteligibilidade (acertos)	Valor de p
<i>keels</i>	0	24	0,00	0,046 <sup>b</sup>
<i>keep</i>	22	2	5,50	0,059 <sup>a</sup>
<i>keepers</i>	13	11	3,25	0,854 <sup>a</sup>
<i>keyboards</i>	20	4	5,00	0,102 <sup>a</sup>
<i>keys</i>	5	19	1,25	0,102 <sup>b</sup>
<i>kiwi</i>	4	20	1,00	0,109 <sup>b</sup>
TOTAL	64	80	16,00	0,854 <sup>a</sup>

QUADRO 25 – Média de inteligibilidade dos lexemas iniciados com segmento oclusivo velar surdo /k/ para a ferramenta de reconhecimento do YouTube.

Comparando o quantitativo de inteligibilidade do fonema em si e dos lexemas iniciados em [k], percebe-se que há uma disparidade de reconhecimento pela ferramenta de transcrição, em que o fonema foi mais reconhecido que o lexema. Isso pode indicar que o segmento oclusivo em *onset* pode não ser fator determinante para a inteligibilidade dos vocábulos, mas que outros fatores linguísticos tenham influenciado a falta de reconhecimento (qualidade da vogal, aspiração da oclusiva inicial, familiaridade lexical, função sintática, etc.).

Estatisticamente, pode-se observar que, para *keep*, como foi apontado em análise do segmento inicial deste vocábulo, o valor de inteligibilidade se mostra marginalmente significativo, aproximando-se de  $p < 0,05$ . O reconhecimento do vocábulo pode estar também relacionado à sua familiaridade ao grupo de falantes e à frequência de uso. Sintaticamente, na estrutura frasal em que foi inserido, a probabilidade de reconhecimento do verbo pode ter sido facilitada devido à configuração básica do *software*, que tenta prever o termo que mais provavelmente ocuparia tal posição sintática baseado no *input* oral que lhe foi apresentado.

Inversamente, a palavra-alvo *keels* não foi inteligível à ferramenta de transcrição. Apesar de os resultados estatísticos não concluírem que a inteligibilidade de [k] em *keels* tenha sido significativa (conforme o Quadro 24),

o fonema [k] apresentou inteligibilidade que se aproximou do valor estatisticamente significativo de  $p$ , tendo sido reconhecido em mais da metade das ocorrências (13 de 24). O teste de Wilcoxon conduzido para o vocábulo mostra que a falta de inteligibilidade deste é estatisticamente significativa ( $p < 0,05$ ) e retoma a ideia de que a familiaridade lexical tenha tido impacto no reconhecimento pela ferramenta de transcrição. O formulário de familiaridade aplicado aos aprendizes brasileiros (grupo de falantes) indicou que o termo não era familiar a nenhum deles. Somado a isso, o *corpus* do COCA indica que a palavra-alvo *keels* não está entre as 5000 mais frequentes no Inglês Americano, o que pode ter feito com que o *software* optasse por outras que atendessem ao contexto fonológico correspondente ao *input* dos falantes.

Em observação aos resultados da transcrição realizada pela ferramenta de reconhecimento de fala do YouTube, percebe-se uma tendência semelhante à que ocorreu com a transcrição de *peach*. No caso de *keels*, o vocábulo foi transcrito em algumas ocorrências como *skills*, de acordo com o que se observa no Quadro 26.

Sujeito	Transcrição do YouTube
01	you got skills to fix stereos
02	I got skills to fit
10	you got skills to fix period
	I got skills to experience
	I got skills to be experienced

QUADRO 26 – Transcrição do YouTube para as frases em que houve a inserção da fricativa [s] anteriormente ao segmento [k] na palavra-alvo *keels*.

Parece haver um encurtamento da vogal que sucede a consoante-alvo (de /i:/ para /ɪ/) que influencia a seleção lexical determinada pela ferramenta de transcrição. Percebe-se também que, somado a isso, o ruído da gravação possa ter determinado a seleção de um vocábulo iniciado em fricativa (/s/).

## 4.2 A INTELIGIBILIDADE PARA O GRUPO DE OUVINTES

Para fins de comparação, amostras de fala dos aprendizes brasileiros também foram submetidas à transcrição de ouvintes humanos. Conforme descrito na Seção 3.2.3 desta dissertação, um recorte das amostras de fala dos

falantes 01, 02, 03 e 10 foi selecionado para a elaboração do experimento de inteligibilidade que foi aplicado aos participantes do grupo de ouvintes: 2 ouvintes brasileiros (01.1 e 02.1) e dois americanos (03.1 e 04.1). Em resumo, os ouvintes foram orientados (por meio de formulário elaborado com a ferramenta *Google Forms*<sup>36</sup>) a ouvir somente uma vez cada amostra de fala e a transcrevê-las. As frases que foram analisadas, a título de comparação com a transcrição do YouTube, são as que se descreve no Quadro 27.

Sujeito	Segmento-alvo	Palavra-alvo	Frase
01	/k/	<i>keepers</i>	<i>We help keepers downstairs.</i>
	/p/	<i>pizza</i>	<i>We eat pizza for lunch.</i>
	/t/	<i>teaspoons</i>	<i>We ship teaspoons for you.</i>
02	/t/	<i>teaser</i>	<i>We talk teaser sentences.</i>
	/p/	<i>peanuts</i>	<i>We ate peanuts yesterday.</i>
	/k/	<i>keep</i>	<i>We keep coffee here.</i>
03	/p/	<i>people</i>	<i>We hit people downstairs.</i>
	/k/	<i>keels</i>	<i>We got keels to fix.</i>
	/t/	<i>teens</i>	<i>You kelp teens at school.</i>
10	/p/	<i>pizza</i>	<i>We eat pizza for lunch.</i>
	/k/	<i>keepers</i>	<i>We help keepers downstairs.</i>
	/t/	<i>teaspoons</i>	<i>We ship teaspoons for you.</i>

QUADRO 27 – Relação das frases transcritas pelos ouvintes referentes às amostras de fala dos falantes 01, 02, 03 e 10.

Com esse recorte, ao todo, então, obteve-se a transcrição de 48 frases (3 segmentos x 4 falantes x 4 ouvintes) feita pelos quatro sujeitos do grupo de ouvintes. O Apêndice H desta dissertação traz as transcrições dos ouvintes referentes às amostras recortadas da coleta com os quatro falantes brasileiros denominados 01, 02, 03 e 10 neste trabalho. Devido ao volume pouco expressivo de dados, não se pôde conduzir testes estatísticos, mas uma análise descritiva das transcrições é trazida, inicialmente, no Quadro 28, dedicado à exposição da inteligibilidade do segmento-alvo para o grupo de ouvintes.

Segmento-alvo	01.1 br <sup>37</sup>	02.1 br	03.1 am <sup>38</sup>	04.1 am	Total
---------------	-----------------------	---------	-----------------------	---------	-------

<sup>36</sup> Google Forms. Disponível em: <<https://docs.google.com/forms/>> Acesso em: 19 Ago. 2017.

<sup>37</sup> Ouvinte brasileiro.

<sup>38</sup> Ouvinte americano.

[k] em <i>keepers</i>	sim	sim	sim	sim	4
[k] em <i>keep</i>	sim	sim	sim	sim	4
[k] em <i>keels</i>	sim	<b>não ouviu</b>	sim	sim	3
[k] em <i>keepers</i>	sim	sim	sim	sim	4
[p] em <i>pizza</i>	sim	sim	sim	sim	4
[p] em <i>peanuts</i>	sim	<b>não ouviu</b>	sim	sim	3
[p] em <i>people</i>	sim	sim	sim	sim	4
[p] em <i>pizza</i>	sim	sim	sim	sim	4
[t] em <i>teaspoons</i>	sim	sim	sim	sim	4
[t] em <i>teaser</i>	sim	<b>não</b>	sim	<b>não</b>	2
[t] em <i>teens</i>	sim	sim	sim	sim	4
[t] em <i>teaspoons</i>	sim	sim	<b>não</b>	sim	3
Total de palavras inteligíveis	12	9	11	11	43

QUADRO 28 - Inteligibilidade do segmento-alvo referente a cada palavra-alvo transcrita pelos quatro ouvintes.

No Quadro 28, indica-se a inteligibilidade dos segmentos-alvo referentes a cada palavra-alvo transcrita para os quatro ouvintes, indicando pela palavra “sim” os momentos em que a consoante-alvo foi reconhecida e por “não” quando não o foi. Ainda, apresenta-se o total de vezes em que cada segmento foi reconhecido em função da palavra-alvo e em função do ouvinte. Houve momentos em que o ouvinte reportou não ter ouvido (ou reconhecido) uma parte da frase a que foi exposto, indicando a situação por um traço (-) na transcrição. Isto é, como cada ouvinte foi orientado a ouvir somente uma vez a cada amostra sonora (simulando o que acontece com a ferramenta de transcrição), quando o sujeito não conseguiu compreender a palavra a ser transcrita, indicou a lacuna na transcrição por um traço. No quadro, essas ocorrências foram indicadas pela etiqueta “não ouviu”.

Observa-se, pelos dados apresentados, que houve alta inteligibilidade de todos os segmentos para os sujeitos ouvintes, em que [k] foi inteligível em 15 das 16 amostras, assim como [p], enquanto [t] foi inteligível em 13 das 16 ocorrências. Essa relação de inteligibilidade também foi observada na transcrição das amostras feita pela ferramenta do YouTube, em que se percebeu ranqueamento similar ([k > p > t]) indicado por teste de Friedman.

Ressalta-se que para [k] e [p], as vezes em não houve inteligibilidade dos segmentos oclusivos estão relacionadas às ocorrências em que o ouvinte



não transcreveu palavra alguma, indicando, na transcrição, com um traço a palavra faltante (*We got (-) to fix.* e *I ate (-) yesterday.*). Em relação ao segmento [t], é possível dizer que ele tenha menor inteligibilidade para o grupo de ouvintes, mas a diferença pode não ser significativa, uma vez que o volume de dados não permite a condução de testes estatísticos para suportar essa afirmação.

Para [t] em *teaser*, parece que houve algum problema de reconhecimento, haja vista que em metade das transcrições o segmento não foi inteligível. A amostra em que o vocábulo *teaser* aparece e que foi selecionada para a elaboração do experimento de inteligibilidade também não foi inteligível à ferramenta de transcrição do YouTube. Isso pode indicar que, para /t/, a ferramenta de transcrição e os ouvintes humanos se comportam da mesma maneira, no sentido de realizar operações de reconhecimento que se assemelham.

É importante, lembrar, entretanto, que a familiaridade lexical pode ter sido determinante tanto para o reconhecimento do segmento, uma vez que a amostra em que *teaser* não foi reconhecida pelo *software* e por dois dos ouvintes não era familiar ao falante. Ainda, o sujeito ouvinte 02.1 também reportou não reconhecer nem saber o que significa o vocábulo.

Para evidenciar a inteligibilidade lexical, o Quadro 29 quantifica a inteligibilidade das palavras-alvo para cada um dos sujeitos do grupo de ouvintes usando o mesmo código do Quadro 28, “sim” para as palavras inteligíveis, “não” para as não inteligíveis e “não ouviu” para os momentos em que o ouvinte indicou por um traço a falta da palavra na transcrição.

Segmento	Palavra-alvo	01.1 br <sup>39</sup>	02.1 br	03.1 am <sup>40</sup>	04.1 am	Total
[k]	<i>keepers</i>	sim	sim	sim	sim	4
	<i>keep</i>	sim	sim	sim	sim	4
	<i>keels</i>	<b>não</b>	<b>não ouviu</b>	<b>não</b>	<b>não</b>	0
	<i>keepers</i>	sim	sim	sim	sim	4
[p]	<i>pizza</i>	sim	sim	sim	sim	4
	<i>peanuts</i>	sim	<b>não ouviu</b>	sim	<b>não</b>	2
	<i>people</i>	sim	sim	sim	sim	4
	<i>pizza</i>	sim	sim	sim	sim	4

<sup>39</sup> Ouvinte brasileiro.

<sup>40</sup> Ouvinte americano.

[t]	<i>teaspoons</i>	sim	<b>não</b>	sim	sim	3
	<i>teaser</i>	sim	<b>não</b>	sim	<b>não</b>	2
	<i>teens</i>	<b>não</b>	sim	sim	<b>não</b>	2
	<i>teaspoons</i>	sim	<b>não</b>	<b>não</b>	sim	2
Total de palavras inteligíveis		10	7	10	8	35

QUADRO 29 - Inteligibilidade de cada palavra-alvo transcrita pelo grupo de quatro ouvintes.

Os resultados somados de inteligibilidade (por ouvinte e por palavra-alvo) mostram que os lexemas aparentam ser menos inteligíveis que o fonema inicial em si. Do total de amostras transcritas, em 43 delas houve o reconhecimento do segmento oclusivo em *onset* e em 35 houve o reconhecimento da palavra-alvo. Isso pode ser um indicativo de que a consoante inicial tenha papel determinante na inteligibilidade lexical para o grupo de ouvintes.

Em comparação ao que se pôde observar na análise da transcrição feita pela ferramenta do YouTube, parece haver uma ordem semelhante de ranqueamento entre os vocábulos-alvo mais inteligíveis e os menos inteligíveis. Os testes de Friedman conduzidos apontaram que, para a ferramenta de transcrição, [p] e [k] são mais inteligíveis que [t], como pode ser, também, observado em relação aos dados coletados com a transcrição do grupo de ouvintes.

Em detalhamento, nota-se que o vocábulo *keels* tem baixa inteligibilidade. Isso foi evidenciado nas amostras da transcrição do YouTube e se repetem para os ouvintes humanos. A amostra de fala recortada para a elaboração do experimento de inteligibilidade em que o vocábulo *keels* foi empregado também não foi reconhecida pela ferramenta de transcrição, assim como não o foi para o grupo de ouvintes. Nesse caso, é possível que a baixa frequência de uso da palavra tenha determinado a sua inteligibilidade, pois a palavra-alvo não foi inteligível nem aos ouvintes humanos nem à ferramenta de transcrição do YouTube em nenhuma de suas ocorrências.

Ainda a qualidade da amostra sonora pode ter causado a falta de inteligibilidade, uma vez que, para a ferramenta de transcrição, a familiaridade lexical não é impeditivo para o reconhecimento, a menos que o vocábulo não esteja incorporado ao seu banco de dados.

Ao observar a transcrição do grupo de ouvintes para a frase em que o vocábulo-alvo *keels* foi empregado, percebe-se um padrão de reconhecimento que se assemelha àquele realizado pelo *software*. O Quadro 30 traz a transcrição da amostra pela ferramenta de reconhecimento de fala do YouTube e apresenta a transcrição dos ouvintes humanos para comparação.

Amostra	Transcrição
Sujeito 01.1	We got kills to fix.
Sujeito 02.1	We got (-) to fix.
Sujeito 03.1	We got kills to fix.
Sujeito 04.1	We got kills to fix. Period.
YouTube	I got killed to fix period

QUADRO 30 – Transcrição da frase com a palavra-alvo *keels* pela ferramenta de transcrição do YouTube e pelo grupo de ouvintes.

Com exceção da amostra do sujeito 02.1 (que não transcreveu o vocábulo), pode-se notar nas demais transcrições que parece ter havido a influência da vogal posterior ao segmento-alvo no reconhecimento do vocábulo. O encurtamento da vogal anterior alta ([ɪ] ao invés de [i:]) pode ter acarretado na seleção de outro exemplar lexical, ocasionando os resultados referentes à inteligibilidade do termo.

Os vocábulos *teaser* e *teens*, ambos iniciados pelo segmento-alvo [t] foram inteligíveis aos ouvintes em metade das amostras. Apesar do volume de dados não ser expressivo, percebeu-se um padrão de reconhecimento nas frases em que as palavras-alvo iniciadas em /t/ foram inseridas. Para *teens*, principalmente, a ferramenta de transcrição acabou realizando um movimento de deslocamento da consoante-alvo para a posição de *coda* do verbo que a precedia. Isso pode ser percebido também na transcrição realizada pelo grupo de ouvintes, registrando-se uma ocorrência para a transcrição de *teens* e duas para a transcrição de *teaser* (*We talked easier sentences.*; *We talked to your sentences. Period.* e *You helped teams at school. Period.*).

Em retomada, percebeu-se a recorrência de dois processos de reconhecimento, tanto para o *software* quanto para o grupo de ouvintes, que podem ter impactado a inteligibilidade dos segmentos e palavras-alvo. O primeiro destacado é o encurtamento da vogal posterior à consoante-alvo, o que fez com que as duas categorias de ouvintes (mecânico e humano)

buscassem alternativas lexicais para a transcrição mais próxima do *input* a que foram expostos. O segundo é o movimento de migração da consoante para o final da palavra que antecedia o vocábulo-alvo.

Nesse processo de comparação das transcrições realizadas pela ferramenta de reconhecimento de fala do YouTube e pelo grupo de ouvintes, destacou-se a pluralidade de experiências linguísticas do grupo de ouvintes. Com a colaboração de dois ouvintes brasileiros e dois ouvintes americanos, pode-se comparar também os dados gerados intragrupo, isto é, comparar a transcrição realizada por cada ouvinte em relação aos demais integrantes do grupo de ouvintes.

Nesse sentido, percebeu-se que não houve disparidade no processo de reconhecimento dos segmentos e das palavras-alvo do experimento quando cruzados os dados dos brasileiros e dos americanos. É possível, inclusive, perceber um equilíbrio nos recursos empregados para desempenhar a tarefa de transcrição. Admite-se que, para afirmações mais generalistas e categóricas, precisaria-se de mais dados, mas pode-se dizer que há uma tendência de equivalência na inteligibilidade dos segmentos oclusivos surdos para os três transcritores em análise (a ferramenta de transcrição do YouTube, os ouvintes brasileiros e os ouvintes americanos).

## 5 CONSIDERAÇÕES FINAIS

O presente trabalho investigou a inteligibilidade das consoantes oclusivas surdas do inglês em posição de ataque silábico produzidas por brasileiros aprendizes de inglês à ferramenta de transcrição automática do YouTube. Para fins de comparação, algumas amostras de fala do grupo de falantes foram expostas à avaliação de ouvintes humanos, a saber, dois ouvintes brasileiros proficientes em inglês e dois ouvintes americanos.

As análises estatísticas e descritivas permitiram perceber que há uma possível ordem de inteligibilidade entre os segmentos oclusivos surdos, em que [k] é mais inteligível que [p], que é mais inteligível que [t]. Na comparação feita entre as consoantes agrupadas em pares ([p] x [t], [p] x [k], e [t] x [k]) não obteve-se valores estatísticos que pudessem se confirmar significativos em relação às amostras transcritas pelo YouTube. Isso pode dar a entender que todas as consoantes parecem obedecer a um padrão de inteligibilidade para a ferramenta. Em consonância, os dados coletados com o grupo de ouvintes apresentou mesma relação de ranqueamento.

A ordem de inteligibilidade identificada para o reconhecimento (para transcritor do YouTube e para o grupo de ouvintes) dos segmentos-alvo foi também verificada em relação aos vocábulos-alvo iniciados em /p/, /t/ e /k/, ou seja, os vocábulos iniciados por [k] são mais inteligíveis que os iniciados por [p], que são mais inteligíveis que os vocábulos iniciados por [t]. Essa comparação permite que se diga que ambos os transcritores (mecânicos e humanos) realizam operações semelhantes de escolha lexical em função da consoante inicial.

Admite-se que, para afirmações mais generalistas e categóricas, precisaria-se de mais dados, mas pode-se dizer que há uma tendência de equivalência na inteligibilidade dos segmentos oclusivos surdos para os três transcritores em análise, a ferramenta de transcrição do YouTube, os ouvintes brasileiros e os ouvintes americanos. Mesmo assim, os dados coletados e as análises conduzidas respondem afirmativamente à pergunta de pesquisa, permitindo que se diga que os segmentos oclusivos surdos produzidos por aprendizes brasileiros são inteligíveis à ferramenta de transcrição automática do YouTube.

Em desmembramento à pergunta de pesquisa, os dados coletados e as análises conduzidas atenderam ao que se pressupunha com a hipótese **a** desta dissertação: “as palavras-alvo produzidas pelos usuários brasileiros em que aparecem os segmentos /p, t, k/ em posição de ataque silábico são menos inteligíveis que o segmento em si”. Em observação qualitativa das transcrições feitas pela ferramenta de reconhecimento do YouTube e pelo grupo de ouvintes, foram identificados alguns fenômenos linguísticos que podem ter corroborado para a diferença de inteligibilidade entre segmento e vocábulo-alvo.

Dadas as limitações de escopo e tempo dessa investigação, não se pode investigar mais a fundo o impacto das amostras de fala para esses movimentos de reconhecimento. Lista-se, no Quadro 31, alguns fenômenos recorrentes, que podem dar margem a futuras discussões e investigações.

Fenômeno linguístico	Total de ocorrências <sup>41</sup>	Palavra-alvo	Exemplos <sup>42</sup>
Houve troca de oclusiva surda por /θ/ em onset.	39 (1) <sup>43</sup>	<i>peace</i>	<i>you got <b>this</b> like River period</i>
		<i>teens</i>	<i>you help <b>things</b> at school period</i>
		<i>team</i>	<i>We stopped <b>them</b> fights. Period.</i>
Houve migração da oclusiva em onset para oclusiva em coda.	25 (4)	<i>team</i>	<i>I <b>stopped</b> in fight period</i>
		<i>teaspoons</i>	<i>we <b>shipped</b> in spoons for you period</i>
		<i>teaser</i>	<i>We talked <b>easier</b> sentences.</i>
Houve alternância de sonoridade na oclusiva surda em onset.	13 (1)	<i>keels</i>	<i>you got <b>guilty</b> six spirit</i>
		<i>peaks</i>	<i>you are <b>big</b> to get a student</i>
		<i>peace</i>	<i>You got <b>big</b> sky and river. Period.</i>
Houve alternância de consoante oclusiva surda em onset.	15 (1)	<i>keepers</i>	<i>you help <b>people</b> downstairs do it</i>
		<i>teens</i>	<i>you have <b>kings</b> at school to it</i>
		<i>teens</i>	<i>You hekp <b>kids</b> at school.</i>

<sup>41</sup> Total com base nas 432 amostras de fala transcritas pelo YouTube.

<sup>42</sup> O terceiro exemplo de cada fenômeno linguístico é referente à transcrição feita pelo grupo de ouvintes.

<sup>43</sup> O valor entre parênteses indica a quantidade de ocorrências observadas na transcrição feita pelo grupo de ouvintes, com base nas 48 amostras transcritas pelo grupo.

Houve inserção de fricativa em <i>onset</i> .	29 (1)	<i>keels</i>	<i>you got <b>skills</b> to fix period</i>
		<i>peach</i>	<i>we ate <b>speech</b> yesterday period</i>
		<i>teaspoons</i>	<i>We keep <b>stones</b> for you.</i>

QUADRO 31 – Relação descritiva dos fenômenos lingüísticos observados nas transcrições realizadas pela ferramenta do YouTube e pelo grupo de ouvintes.

Dada a forma como as ferramentas de transcrição funcionam (buscando padrões probabilísticos de reconhecimento baseado no contexto), acredita-se que haja campo para investigação de outros fatores além dos que foram ilustrados no Quadro 31. Estudos futuros podem encontrar um interessante objeto de estudo na influência que o contexto fonológico apresenta para o reconhecimento de fala automático feita por ferramenta informática. Apesar de ter-se buscado limitar as variáveis intervenientes, percebeu-se que não só o contexto imediato em que um segmento é inserido pode ter implicações para a sua inteligibilidade.

Toma-se como exemplo as situações apresentadas na segunda linha do quadro 31, em que houve migração da oclusiva em *onset* para oclusiva em *coda*. Em análises que considerem os parâmetros acústicos da fala, as evidências coletadas nas transcrições mecânica e humana mostram que pode haver uma diferença no padrão rítmico dos aprendizes brasileiros em comparação ao padrão de falantes nativos americanos. Essa transferência fonológica, de nível suprasegmental, em geral não é levada em consideração por aplicativos de reconhecimento de fala (King et al., 2006), mas pode ser determinante para a eficácia de funcionamento desta tecnologia.

Em resposta à hipótese **b** dessa dissertação, “os segmentos /p, t, k/ em posição de ataque silábico serão mais inteligíveis em palavras-alvo familiares aos usuários brasileiros”, pode-se afirmar que os resultados obtidos na investigação dos segmentos /p/, /t/ e /k/ em posição de ataque silábico a confirmam. Somado a isso, evidenciou-se que a frequência de uso das palavras-alvo possa ser mais impactante à inteligibilidade que a familiaridade dessas ao falante. A consulta feita ao *corpus* COCA mostrou que os itens que tiveram sua inteligibilidade diminuída estavam ranqueados próximos à posição 5000 ou nem apareciam no *corpus*.

Esses resultados podem ter sofrido o impacto da frequência de uso dos vocábulos-alvo. Conforme apresentado no Capítulo 4 desta dissertação, as palavras mais frequentes se mostraram mais inteligíveis, o que pode ser um indicativo de que a frequência de uso lexical seja um fator mais determinante para a inteligibilidade de segmentos oclusivos que a familiaridade aos falantes ou mesmo os valores de VOT. Ressalta-se que a frequência lexical não foi considerada como variável interveniente à época de coleta dos dados, e que sugere-se que o controle de tal variável em futuras pesquisas pode apontar para resultados mais conclusivos sobre a relação inteligibilidade-frequência lexical.

Para finalizar, tem-se que destacar algumas limitações deste estudo de caráter exploratório. Primeiro, frente ao baixo volume de pesquisas que investigassem a inteligibilidade de amostras sonoras à um aplicativo de reconhecimento de fala, a coleta de dados e os próprios instrumentos de coleta de dados foram revisitados diversas vezes para ajustes, inclusive posteriormente a fase de pilotagem. A falta de tratamento acústico na sala onde a coleta de amostras de fala foi feita pode ter impactado a qualidade das amostras, mantendo ruídos que podem ter sido reconhecidos como fricativas.

Ainda sobre as amostras de fala, a escolha do dispositivo de transcrição impôs que os arquivos de audio tivessem que ser convertidos em arquivos de vídeo. Não se percebeu queda da qualidade da amostra, uma vez que a transcrição feita pelo reconhecedor mecânico e pelo grupo de ouvintes apresentou padrões semelhantes. Mesmo assim, seria interessante que, em pesquisas futuras, se atente para essa situação, controlando a variável que pode vir a se mostrar interveniente. Sabe-se, também, que o pequeno número de amostras transcritas não permitiu que mais testes fossem conduzidos e as restrições impostas pela ferramenta de transcrição automática do YouTube implicaram em decisões metodológicas que correspondiam em partes ao que se ambicionava com este trabalho.

Somado a isso, sugere-se que, para futuras investigações, conduza-se a análise acústica das amostras sonoras, o que permitirá identificar padrões na produção dos falantes que podem causar a falta de inteligibilidade a nível segmental e lexical.



## REFERÊNCIAS

ADAMI, André Gustavo. Automatic speech recognition: From the beginning to the portuguese language. In: **The Int. Conf. on Computational Processing of Portuguese (PROPOR)**. Rio Grande do Sul: Porto Alegre. 2010.

ANDRADE, Marco Antônio Rocca de. **Fundamentos de Modelos de Markov Escondidos (HMM)**. Revista Militar de Ciência e Tecnologia, Rio de Janeiro, 01 maio 2000. Disponível em: <[http://rmct.ime.eb.br/arquivos/RMCT\\_2\\_quad\\_2000/fund\\_modelos\\_Markov\\_escondidos.pdf](http://rmct.ime.eb.br/arquivos/RMCT_2_quad_2000/fund_modelos_Markov_escondidos.pdf)> Acesso em: 30 jan. 2017.

ALVES, Mariane A.; SEARA, Izabel. **Produção de plosivas surdas em inglês e português por falantes brasileiros de inglês como língua estrangeira**. In: Anais do CELSUL. Porto Alegre: UFRGS, 2008.

BARBOSA, Plínio A. **Manual de fonética acústica experimental**: aplicações a dados do português. São Paulo: Cortez, 2015.

BARROS, Diana Pessoa de. In: FIORIN, José Luiz (org). A comunicação Humana. **Introdução à Linguística I**: objetos teóricos. 6 ed. São Paulo: Contexto, 2010.

BENT, Tessa; BRADLOW, Ann R. **The interlanguage speech intelligibility benefit**. Journal of the Acoustical Society of America. 2003;114:1600–1610.

BOERSMA, P.; WEENINK, D. (2007) **Praat**: doing phonetics by computer. Disponível em: <<http://www.praat.org>>.

BECKER, Márcia Regina. **A questão da inteligibilidade do inglês como língua franca**. In: ANAIS DO VII CONGRESSO INTERNACIONAL DA ABRALIN. Curitiba: UFPR, 2011. p. 2789-2801.

BECKER, Márcia Regina. **Inteligibilidade da língua inglesa sob o paradigma de língua franca**: percepção de discursos de falantes de diferentes L1s por brasileiros. Tese (Doutorado em Letras), Universidade Federal do Paraná, Curitiba, 2013.

BRUM DE PAULA, M. (Org.). 4º Seminário de Aquisição Fonológica: Resumos e Programação. Pelotas: Ed. da UFPel, 2013. p.26.

CARDOSO, Denise Porto. **Fonologia da Língua Portuguesa**. São Cristóvão: Universidade Federal de Sergipe, CESAD, 2009.

CHO, Taehong; LADEFOGED, Peter. 1999. **Variation and universals in VOT: evidence from 18 languages**. Journal of Phonetics, 27:207-229

CHOMSKY, Noam. **Syntactic structures**. Walter de Gruyter, 2002.

COHEN, D., CRABTREE, B. **Qualitative Research Guidelines Project**. July 2006. Disponível em: <<http://www.qualres.org/HomeMemb-3696.html>> Acesso em: 07 Ago. 2017.

CRISTÓFARO SILVA, Thaís. **Fonética e fonologia do português**: roteiro de estudos e guia de exercícios. 7. ed. - São Paulo : Contexto, 2003.

CRUZ, Neide Cesar. **Inteligibilidade de pronúncia no contexto de inglês como língua internacional**. Revista Intercâmbio, v. XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

CRUZ, Neide Cesar. **Terminologies and definitions in the use of intelligibility: state-of-the-art**. Rev. Brasileira de Lingüística Aplicada, v. 7, n. 1, 2007.

CRUZ, Neide Cesar. **Familiaridade do Ouvinte e Inteligibilidade da Pronúncia de Aprendizes Brasileiros de Inglês**. In: Revista Horizontes de Linguística Aplicada, v.7, n.1, p.88-103, 2008.

CRYSTAL, D. **An encyclopedic dictionary of language and languages**. Oxford: Blackwell Publishers, 1992.

DE GRÈVE, Marcel; VAN PASSEL, Frans. **Linguística e Ensino de Línguas Estrangeiras**. São Paulo: Pioneira, 1975.

DERWING, Tracey M.; MUNRO, Murray J.; CARBONARO, Michael. **Does Popular Speech Recognition Software Work with ESL Speech?** In: TESOL Quarterly, v. 34, n. 3, 592- 603, 2000.

DERWING, Tracey M.; MUNRO, Murray J. Putting accent in its place: Rethinking obstacles to communication. **Language teaching**, v. 42, n. 4, p. 476-490, 2009.

GILNER, Leah, MORALES, Franc. **Pronunciation: Intelligibility, Frequency Considerations, and Instruction**. 文京学院大学外国語学部文京学院短期大学紀要 9: 35-56, 2010.

JENKINS, Jennifer. **The Phonology of English as an International Language**: New Models, New Norms, New Goals. Oxford: Oxford University Press, 2000.

JENKINS, Jennifer. **A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language**. Applied Linguistics 23/1: 83-103, 2002.

KENT, Ray D.; READ, Charles. **Análise acústica da fala**. 1.ed. São Paulo: Cortez, 2015.

KENWORTHY, Joanne. **Teaching English Pronunciation**. London: Longman, 1987.

KING, Simon; FRANKEL, Joe; LIVESCU, Karen; MCDERMOTT, Erik; RICHMOND, Korin; WESTER, Mirjan. **Speech production knowledge in automatic speech recognition**. Journal of the Acoustical Society of America, 2006.

LAMPRECHT, Regina Ritter; et al. **Aquisição Fonológica do Português: Perfil de Desenvolvimento e Subsídios para Terapia**. Porto Alegre: Artmed, 2004.

LAVER, John. **Principles of Phonetics**. Cambridge: University Press, 1994.

LISKER, Leigh; ABRAMSON, Arthur. (1964). **A Cross-language study of voicing in initial stops: Acoustical measurements**. Word, 20, 384-422.

LOUZADA, Jailton Alkimin. **Reconhecimento Automático de Fala por Computador**. Trabalho de conclusão de curso: Pontifícia Universidade Católica de Goiás, 2010.

MARTINS, Valéria Farinazzo, BRASILIANO, André. **Interface do usuário baseada em voz como ferramenta para promover o ensino/aprendizagem de língua estrangeira**. 2012. Disponível em: <<http://www.revistas.udesc.br/index.php/reavi/article/view/2524>> Acesso em: 18 jan. 2017.

MUSSALIN, Fernanda; BENTES, Anna Christina (org). **Introdução à Linguística I: domínios e fronteiras**. São Paulo: Cortez, 2001.

MUNRO, M. J. **Foreign accent and speech intelligibility**. Phonology and second language acquisition, 193-218, 2008.

OLIVEIRA, Luiz Eduardo Soares de, MORITA, Marisa Emika. **Introdução aos Modelos Escondidos de Markov (HMM)**. Pontifícia Universidade Católica do Paraná – PUC-Pr, 2005.

OTHERO, G.; MENUZZI, S. **Linguística Computacional: teoria e prática**. São Paulo: Parábola Editorial, 2005.

PAIVA, Vera Lúcia Menezes de Oliveira. **Aquisição de Segunda Língua**. São Paulo: Parábola, 2014.

PÉPIOT, Erwan. **Voice, speech and gender**:. Male-female acoustic differences and cross-language variation in English and French speakers. Corela. Cognition, représentation, langage, n. HS-16, 2015.

PEROZZO, Reiner Vinicius; ALVES, Ubiratã Kickhöfel. **Identificação E Discriminação De Oclusivas Não Vozeadas Sem Soltura De Ar Audível Do Inglês: Dados De Aprendizes Brasileiros**. Fórum Linguístico, Florianópolis, v. 11, n. 1, p. 23-37, jan./mar. 2014.

PETTER, Margarida. Linguagem, língua, linguística. In: FIORIN, José Luiz (org.). **Introdução à Linguística I: objetos teóricos**. 6 ed. São Paulo: Contexto, 2010.

MADHU SUDARSHAN REDDY, B.; MAHENDRA KUMAR, N.; SREEDEVI, N. **Voice Onset Time across gender and different vowel contexts in Telugu**. In: *Language in India*, 2014.

SANTOS, Raquel Santana; SOUZA, Paulo Chagas de. Fonética. In: **Introdução à Linguística II: Princípios de Análise**. 5ed. São Paulo: Contexto, 2011.

SAUSSURE, Ferdinand De, Wade Baskin, and Perry Meisel. **Course in general linguistics**. Columbia University Press, 2011.

SCHWARTZHAUPT, B., Alves, U. K., & FONTES, A. **O VOT como pista suficiente para a distinção surdo/sonoro**: dados de falantes do inglês americano. 4º Seminário de Aquisição Fonológica: Resumos e Programação. Pelotas: Ed. da UFPel, 26, 2013.

SCHWARTZHAUPT, Bruno Moraes; ALVES, Ubiratã Kickhöfel. **A influência do contexto fonético-fonológico nos valores de Voice Onset Time**: verificação de três sistemas linguísticos. *Fórum Linguístico*, Florianópolis, v. 11, n.1, p.51-68, jan-mar. 2014.

SMITH, L. E.; NELSON, C.L. International intelligibility of English: directions and resources. **World Englishes** 4, p. 333–342, 1985.

SOUZA, Aline de. **A produção das oclusivas desvozeadas do inglês por aprendizes brasileiros**: uma análise acústica. Trabalho de Conclusão de Curso (Letras), Universidade Tecnológica Federal do Paraná, Curitiba, 2012.

STATISTA. **The most spoken languages worldwide (speakers and native speaker in millions)**. Disponível em: <<http://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/>> Acesso em: 05 set. 2016.

WHITESIDE, S. P., HENRY, L., & DOBBIN, R. Sex differences in voice onset time: a developmental study of phonetic context effects in British English. **Journal of the Acoustical Society of America**, 116, 1179-1183, 2004.

XU, Yi. **In defense of lab speech**. *Journal of Phonetics* 38: 329-336, 2010.

YAVAS, M.; WILDERMUTH, R. **The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers**. *IRAL*, Heidelberg, n.44, p.251-263, 2006.

ZIMMER, M. C.; ALVES, U. K. **Uma visão dinâmica da produção da fala em L2: o caso da Dessonorização Terminal**. *Revista da Abralín*, Brasília, v.11, n.1, p.221- 272, 2012.



## APÊNDICE A – TERMOS DE CONSENTIMENTO LIVRE E ESCLARECIDO

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO - ESCOLA

- a) Você está sendo convidado a participar de um estudo intitulado “A produção de oclusivas surdas em língua inglesa: a inteligibilidade de aprendizes brasileiros e o reconhecimento de fala”. É através das pesquisas científicas que ocorrem avanços importantes em todas as áreas, e sua participação é fundamental.
- b) O objetivo desta pesquisa é investigar a inteligibilidade de falantes brasileiros de inglês como língua estrangeira quando interagem com um aplicativo de reconhecimento de fala.
- c) Caso você participe da pesquisa, será necessário que autorize o acesso aos estudantes da instituição em horário que não atrapalhe os encaminhamentos pedagógicos institucionais para a coleta de dados biográficos e de produção oral.
- d) A todos os sujeitos envolvidos será garantido o sigilo quanto à sua identidade.
- e) Os benefícios esperados como resultados desta pesquisa são:
- e1) Contribuir para melhorar a tecnologia de reconhecimento de voz;
- e2) Divulgar a necessidade de estudos que tenham por foco o estudante de língua estrangeira.
- f) A pesquisadora Anna Cristina Baratieri, Mestranda em Letras, endereço eletrônico [annabaratieri@outlook.com](mailto:annabaratieri@outlook.com) e telefone (41) 9 9973 9582, pode ser contatada por telefone ou endereço eletrônico a qualquer momento e é a responsável pela pesquisa, esclarecendo eventuais dúvidas a respeito desta investigação.
- g) Estão garantidas todas as informações que você queira, antes, durante e depois do estudo.
- h) A sua participação neste estudo é voluntária.
- i) Qualquer informação que for divulgada em relatório ou publicação será feito sob forma codificada, para que a confidencialidade dos participantes seja mantida.
- j) Os dados dos participantes serão gravados, respeitando-se completamente o anonimato. Os dados gerados serão mantidos codificados após o término da pesquisa para fim de composição de *corpus* de pesquisa para eventuais investigações posteriores.
- k) Todas as despesas necessárias para a realização da pesquisa, como material de consumo, papéis, fitas e outros não são da sua responsabilidade.
- l) Pela participação no estudo, ninguém receberá qualquer valor em dinheiro.

**Eu li o texto acima e compreendi a natureza e objetivo da pesquisa para a qual fui convidado a participar. A explicação que recebi menciona os benefícios da pesquisa. Eu concordo voluntariamente em participar desta pesquisa e entendo que a participação é de minha inteira responsabilidade, declarando que os dados informados são reais e que tenho ciência de que não serão individualmente revelados.**

(Nome do responsável legal pela instituição)

---

(Assinatura do responsável legal pela  
instituição)

Curitiba, \_\_\_\_/\_\_\_\_/\_\_\_\_

## **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO - PARTICIPANTE**

a) Você está sendo convidado a participar de um estudo intitulado “A produção de oclusivas surdas em língua inglesa: a inteligibilidade de aprendizes brasileiros e o reconhecimento de fala”. É através das pesquisas científicas que ocorrem avanços importantes em todas as áreas, e sua participação é fundamental.

b) A pesquisadora Anna Cristina Baratieri, endereço eletrônico [annabaratieri@outlook.com](mailto:annabaratieri@outlook.com) e telefone (41) 9 9973 9582, Mestranda em Letras pelo Programa de Pós Graduação em Letras da UFPR, sob orientação da Profª Drª Denise Cristina Kluge, poderá ser contatada por telefone ou endereço eletrônico a qualquer momento e é a responsável pela pesquisa e poderá esclarecer eventuais dúvidas a respeito desta pesquisa.

c) O objetivo desta pesquisa é investigar a inteligibilidade de falantes brasileiros de inglês como língua estrangeira quando interagem com um aplicativo de reconhecimento de fala.

d) Caso você participe da pesquisa, será necessário que responda a um questionário de informações pessoais, que participe de testes gravados no formato de ditados e que interaja com um software de reconhecimento de voz.

e) A todos os sujeitos envolvidos será garantido o sigilo quanto à sua identidade.

f) Os benefícios esperados como resultados desta pesquisa são: 1) contribuir para melhorar a tecnologia de reconhecimento de voz; e 2) divulgar a necessidade de estudos que tenham por foco o estudante de língua estrangeira.

g) Estão garantidas todas as informações que você queira, antes, durante e depois do estudo.

h) A sua participação neste estudo é voluntária. Você tem a liberdade de se recusar a participar ou, se aceitar participar, retirar seu consentimento a qualquer momento.

i) Qualquer informação que for divulgada em relatório ou publicação será feito sob forma codificada, para que a confidencialidade seja mantida.

j) Os seus dados serão gravados, respeitando-se completamente o seu anonimato. Os dados gerados serão mantidos codificados após o término da pesquisa para fim de composição de *corpus* de pesquisa para eventuais investigações posteriores.

k) Todas as despesas necessárias para a realização da pesquisa, como material de consumo, papéis, fitas e outros não são da sua responsabilidade.

l) Pela sua participação no estudo, você não receberá qualquer valor em dinheiro.

**Eu li o texto acima e compreendi a natureza e objetivo da pesquisa para a qual fui convidado a participar. A explicação que recebi menciona os**

benefícios da pesquisa. Eu entendi que sou livre para interromper minha participação na pesquisa a qualquer momento sem justificar minha decisão. Eu concordo voluntariamente em participar desta pesquisa e entendo que a participação é de minha inteira responsabilidade, declarando que os dados informados são reais e que tenho ciência de que não serão individualmente revelados.



## TERM OF FREE AND INFORMED CONSENT - PARTICIPANT

- a) You are being invited to participate in a study called “The voiceless stops in English: intelligibility of Brazilian learners and speech recognition”. It is through scientific research that important development happens in all areas, and your participation is fundamental.
- b) The researcher Anna Cristina Baratieri, graduate student in Letters by UFPR, tutored by Professor Dr. Denise Cristina Kluge, whose e-mail is [annabarateri@outlook.com](mailto:annabarateri@outlook.com) and telephone is +55 41 9 9973 9582, can be contacted at any time and is responsible for the research and can answer to eventual questions about this investigation.
- c) The goal of this research is to investigate the intelligibility of Brazilian speakers of English as Second Language when they interact with speech recognition apps.
- d) If you take part in this study, it will be asked you to fill out a biographical questionnaire, to take part in recorded experiments in a dictation form and to interact with speech recognition software.
- e) To all the involved subjects it will be guaranteed secrecy about their identity.
- f) The expected benefits as results from this investigation are:
  - f1) Contribute to improve voice recognition technology;
  - f2) Publicize the need of studies that aim for the foreign language student.
- g) It is guaranteed all the information you may request, before, during, and after the study.
- h) This is a volunteer participation. You have the right to refuse to participate or, if you agree, withdraw your consent at any time.
- i) Any information that may be released in report or publication will be at coding form, to assure the confidentiality of participant data.
- j) Your data will be recorded, fully respecting your anonymity. The data generated will be kept coded after the investigation period with the intention of creating research *corpus* for eventual future researches.
- k) All expenses due to the research, such as paper, tape and recording material, are not of your responsibility.
- l) For your participation in this study, you will not receive any financial reward.

**I read the text above and understood the nature of the research which I have been invited to participate. The explanation I received mentions the benefits of this study. I understood that I am free to withdraw my participation in this research at any moment without having to explain my decision. I voluntarily agree to participate in this investigation and I understand that my participation is my entire responsibility, declaring that the informed data are real and that I acknowledge that they will not be individually released.**

**APÊNDICE B – QUESTIONÁRIO BIOGRÁFICO PARA BRASILEIROS****Programa de Pós-Graduação em Letras****Área de Concentração: Estudos Linguísticos****Mestranda: Anna Cristina Baratieri****Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Denise C. Kluge****Questionário 101****A PRODUÇÃO DE OCLUSIVAS SURDAS EM LÍNGUA INGLESA: A INTELIGIBILIDADE DE APRENDIZES BRASILEIROS E O RECONHECIMENTO DE FALA**

Gostaríamos de solicitar sua ajuda respondendo às questões que seguem relativas à sua experiência com a Língua Inglesa. Esta pesquisa é conduzida por Anna Cristina Baratieri, como parte de seu projeto de mestrado em letras pela UFPR, sob orientação da Prof<sup>a</sup>. Dr<sup>a</sup>. Denise C. Kluge, e será útil para melhorar nosso entendimento sobre a inteligibilidade de falantes brasileiros de Inglês como Língua Estrangeira na interação com um software de reconhecimento de voz. As suas respostas às questões serão mantidas sob estrita confidencialidade. Nosso interesse recai sobre sua experiência pessoal. Por favor, responda com sinceridade, haja vista que somente assim poderemos garantir o sucesso da investigação. Obrigada por sua colaboração.

**PARTE I**

Por favor, responda a este questionário assinalando a resposta que melhor se aplica a sua realidade e/ou respondendo nos espaços fornecidos. Os dados gerados neste estudo serão agregados e os resultados publicados manterão a confidencialidade das respostas individuais.

1. Qual é a sua língua materna?

---

2. Em qual idioma você foi alfabetizado?

---

3. Você fala outros idiomas? Se sim, Qual(is)?

---

4. Você já estudou inglês em escolas de idiomas?

( ) sim

( ) não

4.1 Se sim, qual(is) escola(s)?

---

4.2 Por quanto tempo?

---

4.3. Com que idade você começou a estudar inglês em escola de idiomas?

---

5. Você ainda estuda em escolas de idiomas?

( ) sim                      ( ) não

5.1 Se sim, qual?

---

5.2 Há quantos semestres você é aluno desta escola?

---

5.3 Quantas aulas semanais você tem?

---

5.4 Qual é a duração das aulas?

---

5.5 Qual é variedade do inglês você considera que fala?

( ) africana

( ) americana

( ) australiana

( ) britânica

( ) outro: \_\_\_\_\_

6. Com que idade você começou a estudar inglês na escola regular? (Ensino Fundamental e Médio)

---

7. Você já morou em algum outro país que não seja o Brasil?

( ) sim                      ( ) não

7.1 Se sim, qual(is)?

---

7.2 Por qual período (mês/ano – mês/ano)?

---

8. Você já viajou a passeio ou negócios para algum país de língua inglesa?

( ) sim                      ( ) não

8.1 Se sim, qual?

---

8.2 Por quanto tempo (em dias)?

---

9. Você já participou de algum curso ou teve aulas de pronúncia de Inglês?

( ) sim                      ( ) não

9.1 Se sim, por qual período (mês/ano – mês/ano)?

---

10. Você já fez algum teste de proficiência de Inglês?

( ) sim                      ( ) não

10.1 Se sim, qual?

---

10.2 Quando (ano)?

---

10.3 Caso lembre, qual foi seu resultado?

---

11. Você tem algum problema auditivo diagnosticado? Se sim, qual?

---

12. Você tem algum problema de visão diagnosticado? Se sim, qual?

---

13. Você tem algum comprometimento de fala diagnosticada? Se sim, qual?

---

14. Complemente com qualquer informação que você considera relevante sobre o seu contato com a língua inglesa.

---

---

---

## PARTE II

O objetivo deste questionário biográfico é apenas o de auxiliar na análise de dados de pesquisa de mestrado na área de fonética e fonologia da língua inglesa. Os nomes dos participantes não serão, sob hipótese alguma, divulgados.

**Nome completo**



## APÊNDICE C – ROTEIRO DE ENTREVISTA

### ROTEIRO DE ENTREVISTA

#### I. Abertura

- A. Olá, meu nome é Anna Baratieri e como mestranda em Letras pela UFPR, pensei que seria uma boa ideia entrevistar você antes de dar continuidade aos experimentos que vamos conduzir.
- B. Eu gostaria de fazer algumas perguntas gerais sobre as respostas que foram dadas no questionário biográfico, para que não haja dúvidas quando a sua intenção de resposta.
- C. Eu espero usar essas informações para auxiliar na interpretação dos dados coletados nas próximas etapas da pesquisa
- D. A entrevista e os experimentos consecutivos não devem demorar mais que vinte minutos. Você tem essa disponibilidade de tempo para isso agora?

Transição: Deixe-me começar perguntando-lhe algumas coisas sobre você e sua experiência.

#### II. Corpo

##### A. Informação linguística geral

- 1. Há quanto tempo você estuda Inglês?
  - a. Você sente que tem facilidade no aprendizado de idiomas?
  - b. Eu li em seu questionário que você fala outro idioma além do Inglês. Como você descreveria o seu nível de fluência nesse idioma?
- 2. Eu li em seu perfil que você já morou em outro país.
  - a. Como foi essa experiência?
  - b. Qual foi o propósito desta viagem?
  - c. Você acha que o seu desempenho com o idioma melhorou? Se sim, de que maneira?

(Transição para o próximo tópico)

##### B. Educação

- 1. Eu li no questionário que você respondeu que já participou de aulas de pronúncia.
  - a. Onde você teve estas aulas?
  - b. Como era a estrutura da aula?
  - c. Você recomendaria as aulas de pronúncia?

(Transição para o próximo tópico)

##### C. Atividades/Hobbies/Interesses

- 1. Durante o dia, você utiliza tecnologia para se comunicar?
  - a. Quais formas de tecnologia você se sente mais confortável usando?
  - b. Você acha que esses aplicativos facilitam e aceleram o tempo de comunicação?

- c. Você se sente confortável utilizando tecnologia?
- 2. Você sabe que esta pesquisa trata do uso de aplicativos de reconhecimento de fala?
  - a. Você já utilizou algum para dar comandos ou ditar textos?
  - b. Se sim, como foi sua experiência?

(Transição para o próximo tópico: bom, foi um prazer descobrir mais coisas sobre você. Deixe-me resumir as informações que foram gravadas durante a entrevista.)

### **III. Fechamento**

- A. Você está realmente envolvido(a) com tecnologia no seu dia-a-dia. Você já teve experiências em \_\_\_\_\_. Seus contatos são feitos geralmente via \_\_\_\_\_.
- B. Eu agradeço ter disposto este tempo para a entrevista. Tem alguma coisa a mais que você acha que poderia ser útil que eu saiba para que a análise dos dados coletados possa ser feita da melhor forma possível?
- C. Eu devo ter toda a informação necessária agora. Obrigada mais uma vez. Espero poder contar com sua participação em investigações futuras.

## APÊNDICE D – QUESTIONÁRIO DE FAMILIARIDADE

### Questionário de Familiaridade

Por favor, responda a este questionário assinalando a resposta que melhor se aplica a sua realidade. Os dados gerados e os resultados publicados manterão a confidencialidade das respostas individuais. Com a primeira pergunta, pretende-se investigar se o participante já viu ou ouviu a palavra, mesmo que não consiga empregá-la em contexto comunicacional. Com a segunda, se o participante conhece o significado da palavra.

#### **PENCIL**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PEACE**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PEACH**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PIZZA**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PEANUT**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PEOPLE**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **PEAK**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **POOR**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEETH**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEEN**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEAM**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TABLE**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEACHER**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEASPOON**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TEASER**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **TOOL**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KEY**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KEEP**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KIWI**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **COOK**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KEEPER**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KEYBOARD**

Você conhece esta palavra?  
 Você sabe o seu significado?

#### **KEEL**


Você conhece esta palavra?  
 Você sabe o seu significado?


#### **KING**


Você conhece esta palavra?  
 Você sabe o seu significado?




## APÊNDICE E – CARTÕES EMPREGADOS NO EXPERIMENTO DE FAMILIARIDADE

PEACH
A round fruit with soft red and yellow skin, yellow flesh and a large rough seed inside


PIECE
An amount of something that has been cut or separated from the rest of it


PEAK
The pointed top of a mountain; a mountain with a pointed top


PIZZA
An Italian dish consisting of a flat round bread base with cheese, tomatoes, vegetables, meat, etc. on top


**PEANUT**

A nut that grows underground in a thin shell

**PEOPLE**

Persons; men, women and children

**KEY**

A specially shaped piece of metal used for locking a door, starting a car, etc

**KEEP**

To stay in a particular condition or position; to make somebody/something do this



**KEEL**

The long piece of wood or steel along the bottom of a ship, on which the frame is built, and which sometimes sticks out below the bottom and helps to keep it in a vertical position in the water

**KIWI**

A small fruit with thin brown skin covered with small hairs, soft green flesh and black seeds

**KEYBOARD**

The set of keys for operating a computer or typewriter

**KEEPER**

A player whose job is to stop the ball from going into his or her own team's goal



**TEETH**

Any of the hard white structures in the mouth used for biting and chewing food

**TEEN**

Between 13 and 19 years old; connected with people of this age

**TEAM**

A group of people who play a particular game or sport against another group of people

**TEACHER**

A person whose job is teaching, especially in a school



**TEASPOON**

A small spoon for putting sugar into tea and other drinks

**TEASER**

An advertisement for a product that does not mention the name of the product or say much about it but is intended to make people interested and likely to pay attention to later advertisements





## APÊNDICE F – LÂMINAS DE ORIENTAÇÃO PARA A COLETA DE AMOSTRAS SONORAS

### ORIENTAÇÕES AO PARTICIPANTE

- A seguir, você fará a leitura de frases em língua inglesa.
- Antes que o experimento inicie, faremos uma prática, para que entenda como funciona.
- Lembre-se que não podemos parar o experimento. Assim, ele só se encerra assim que as 108 frases forem lidas.
- A cada 10 frases, você receberá uma notificação sobre o avanço do experimento.
- Ao final de cada frase, diga PERIOD, para que o software saiba que uma nova frase se inicia.

LÂMINA 1

### Vamos tentar?

- As próximas quatro frases estão formatadas conforme o padrão do experimento.
- Leia-as em voz alta e lembre-se de dizer PERIOD ao final de cada uma.
- Assim que terminar de lê-las faremos uma última checagem de orientação e daremos início ao experimento.
- Se ainda persistir alguma dúvida sobre como operar, diga TESTE e volte às quatro frases de adaptação que seguem para treinar mais um pouco.

LÂMINA 2

## E aí?

- Agora que o formato do experimento lhe é familiar, vamos ao *checklist*:
  - O microfone está ajustado confortavelmente? Se não, fique à vontade para ajustá-lo.
  - Você lembrou de dizer PERIOD após cada frase?
- **Você pode repetir as frases se achar que não pronunciou-as com clareza. Neste caso, lembre-se de dizer PERIOD ao final da repetição.**

LÂMINA 7

## Vamos começar?

- A partir da próxima tela, os dados gravados serão utilizados para a análise proposta por esta pesquisa.
- Lembre-se que não podemos fazer pausas no experimento.
- Leia as frases em voz alta e lembre-se de dizer PERIOD ao final de cada uma.
- Obrigada, mais uma vez, e **DIVIRTA-SE!**

LÂMINA 8

## APÊNDICE G – TUTORIAL PARA TRANSCRIÇÃO DE VÍDEOS NO YOUTUBE

1. Se seu arquivo já for um arquivo de vídeo, o processo é simplificado e você já pode importar o seu vídeo para o YouTube. Se for um arquivo de áudio, você terá que transformá-lo em vídeo primeiro.
2. Você tem que ter uma conta no Google (*e-mail*) para fazer a importação do arquivo de vídeo para o YouTube, bem como um canal no YouTube.
3. Ao fazer a importação do vídeo, preencha os campos solicitados (nome, etiquetas, descrição, etc.) e altere a visualização do vídeo de PÚBLICO para PARTICULAR (último campo), caso não queira que outras pessoas vejam seu vídeo. Feito isso, clique em Publicar.
4. Quando o navegador tiver concluído a importação do vídeo, clique sobre o campo onde aparece a foto do usuário no canto superior direito da página de navegação e, depois, no *link* Creator Studio.
5. Localize o vídeo que pretende transcrever e clique na seta ao lado da função *Edit* do vídeo. A barra de rolagem permitirá que você selecione a indicação *Escolher Subtitles & CC*
6. Escolha ou altere a o idioma de transcrição para o que for correspondente ao áudio do arquivo que você importou clicando em Change Language e aguarde (pelo menos 10 minutos). Você perceberá uma barra acinzentada crescente na margem inferior da janela de vídeo. É fundamental que você não abandone o navegador ou visite outras páginas enquanto a barra acinzentada não estiver completamente preenchida, ou a transcrição será interrompida.
7. Por fim, clique no *link* Vídeos, em *Edit* e em *Subtitles*. Lá você poderá ter acesso às transcrições e legendagens feitas para o vídeo. Clique na que você gerou que o navegador te redirecionará para a página de edição de legendas.
8. Você pode editar a legenda de acordo com a amostra sonora do vídeo ou mantê-la conforme foi transcrito pela ferramenta de transcrição automática.
9. Ao final, é possível salvar o arquivo de legenda em formato .txt para edição ou análise posterior.



**APÊNDICE H – TRANSCRIÇÃO DAS AMOSTRAS DOS SUJEITOS 01, 02, 03 E 10 FEITA PELOS OUVINTES**

<b>Amostra</b>	<b>01.1</b>	<b>02.1</b>	<b>03.1</b>	<b>04.1</b>
1	I help "keepers" downstairs.	I help keepers downstairs.	I help keepers downstairs.	I helped keepers downstairs.
2	I eat pizza for lunch.	I eat pizza for lunch.	I eat pizza for lunch.	I eat pizza for lunch.
3	We ship teaspoons for you.	We (-) tissues for you.	We ship teaspoons for you.	We sheeped teaspoons for you.
4	We talk teasers sentences.	We talked easier sentences.	We talk teaser sentences.	We talked to your sentences.
5	I ate peanuts yesterday.	I ate (-) yesterday.	I ate peanuts yesterday.	I ate pianos yesterday.
6	I keep coffee here.	I keep coffee hear.	I keep coffee here.	I keep coffee here.
7	We - people downstairs.	We (-) people downstairs.	We get people downstairs.	We hid people downstairs.
8	We got kills to fix.	We got (-) to fix.	We got kills to fix.	We got kills to fix.
9	You hekp kids at school.	You help teens at school.	You help teens at school.	You helped teams at school.
19	You ate pizza for lunch.	You ate pizza for lunch.	You ate pizza for lunch.	You eat pizza for lunch.
20	You help keepers downstairs.	You help keepers downstairs.	You help keepers downstairs.	You helped keepers downstairs.
21	We ship teaspoons for you.	We (-) tissues for you.	We keep stones for you.	We shipped teaspoons for you.

**APÊNDICE I – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 01 FEITA  
PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE**

<b>Palavra-alvo</b>	<b>Item</b>	<b>Transcrição youtube</b>	<b>Frase veículo</b>
keels	17	i got cubes to fix period	I got keels to fix.
keels	27	I got you six period	I got keels to fix.
keels	35	you got skills to fix stereos	You got keels to fix.
keels	62	we got killed fifth period	We got keels to fix.
keels	69	you got Q to face the real	You got keels to fix.
keels	98	we got to fix period	We got keels to fix.
keep	3	we keep cocked here to use	We keep coffee here.
keep	16	I keep coffee here period	I keep coffee here.
keep	34	you keep coffee here period	You keep coffee here.
keep	42	I keep coughs here period	I keep coffee here.
keep	56	you keep coffee here period	You keep coffee here.
keep	85	we keep caught here period	We keep coffee here.
keepers	2	you help keepers downstairs to use	You help keepers downstairs.
keepers	7	you help keepers downstairs period	You help keepers downstairs.
keepers	21	I help keepers downstairs period	I help keepers downstairs.
keepers	28	I helped keepers downstairs period	I help keepers downstairs.
keepers	50	we help people downstairs term use	We help keepers downstairs.
keepers	100	we helped keepers downstairs period	We help keepers downstairs.
keyboards	1	we got keyboards to fix periods	We got keyboards to fix.
keyboards	25	you got keyboards to fit the reason	You got keyboards to fix.
keyboards	48	we got keyboards to fix period	We got keyboards to fix.
keyboards	52	I've got keyboards to face commune	I got keyboards to fix.
keyboards	77	you got keyboards to face period	You got keyboards to fix.
keyboards	90	I got keyboard to fix period	I got keyboards to fix.
keys	4	I got keep downstairs period	I got keys downstairs.
keys	30	without keys downstairs period	We got keys downstairs.
keys	63	we got killed downstairs to read	We got keys downstairs.
keys	88	you got to keep the spirit curious	You got keys downstairs.
keys	91	you got to keep downstairs period	You got keys downstairs.
keys	106	I got to keep downstairs to read	I got keys downstairs.
kiwi	22	I eat chili for lunch period	I eat kiwi for lunch.
kiwi	72	I eat to eat for lunch period	I eat kiwi for lunch.
kiwi	74	you eat Kiwi for lunch periods	You eat kiwi for lunch.
kiwi	80	with each Kiwi for lunch period	We eat kiwi for lunch.
kiwi	96	you eat chili for lunch period	You eat kiwi for lunch.
kiwi	104	we eat Kiwi for lunch period	We eat kiwi for lunch.
peace	6	I got peace like a river period	I got peace like a river.

peace	11	you got peace like a river period	You got peace like a river.
peace	24	we got peace like a river period	We got peace like a river.
peace	29	we got peace like a river period	We got peace like a river.
peace	66	I got peace like a river period	I got peace like a river.
peace	68	you got peace like a river - you'd	You got peace like a river.
peach	20	I ate speech yesterday period	I ate peach yesterday.
peach	60	I eat speech yesterday period	I ate peach yesterday.
peach	84	with 8 speech yesterday period	We ate peach yesterday.
peach	87	you ate speech yesterday period	You ate peach yesterday.
peach	94	you ate speech yesterday two years	You ate peach yesterday.
peach	108	we make speech yesterday to	We ate peach yesterday.
peaks	12	we walk pigs together period	We walk peaks together.
peaks	19	you walk pigs together to use	You walk peaks together.
peaks	51	I walk it together to use	I walk peaks together.
peaks	86	i walk pigs together period	I walk peaks together.
peaks	93	we all win war pigs together to you	We walk peaks together.
peaks	95	you walk pigs together to use	You walk peaks together.
peanuts	14	your weights peanuts yesterday period	You ate peanuts yesterday.
peanuts	32	you ate peanuts yesterday period	You ate peanuts yesterday.
peanuts	75	I eat peanuts yesterday period	I ate peanuts yesterday.
peanuts	78	we eat in it yesterday period	We ate peanuts yesterday.
peanuts	99	we eat spinach yesterday period	We ate peanuts yesterday.
peanuts	103	I ate penis yesterday period	I ate peanuts yesterday.
people	39	you heat people downstairs period	You hit people downstairs.
people	53	you need people downstairs period	You hit people downstairs.
people	54	we heat people downstairs period	We hit people downstairs.
people	58	I need people downstairs period	I hit people downstairs.
people	64	we heat people downstairs period	We hit people downstairs.
people	71	I hit people downstairs period	I hit people downstairs.
pizza	10	we eat pizza for lunch period	We eat pizza for lunch.
pizza	40	we eat pizza for lunch period	We eat pizza for lunch.
pizza	44	I eat pizza for lunch period	I eat pizza for lunch.
pizza	45	I eat pizza for lunch period	I eat pizza for lunch.
pizza	55	you eat pizza for lunch period	You eat pizza for lunch.
pizza	89	you eat pizza for lunch period	You eat pizza for lunch.
teachers	8	you help teachers at school period	You help teachers at school.
teachers	18	you helped teachers at school period	You help teachers at school.
teachers	49	I have teachers at school period	I help teachers at school.
teachers	79	we health teachers at school period	We help teachers at school.
teachers	82	I help teachers at school period	I help teachers at school.

teachers	107	the health teachers at school period	We help teachers at school.
team	5	I stopped in fight period	I stop team fights.
team	41	I stopped a fight period	I stop team fights.
team	43	we stop same price period	We stop team fights.
team	61	we saw 10 fights period	We stop team fights.
team	67	you stopped in fight period	You stop team fights.
team	83	you stopped in fight career	You stop team fights.
teaser	9	you talk to your sentences peruse	You talk teaser sentences.
teaser	37	I talked to the sentences period	I talk teaser sentences.
teaser	47	you talk to the sentences period	You talk teaser sentences.
teaser	57	I talked to your sentences period	I talk teaser sentences.
teaser	92	we talked to the sentences career	We talk teaser sentences.
teaser	97	we talk teaser sentences period	We talk teaser sentences.
teaspoons	13	you ship teaspoons for you to peruse	You ship teaspoons for you.
teaspoons	15	we ship systems for you period	We ship teaspoons for you.
teaspoons	38	we shit kiss them for you the real	We ship teaspoons for you.
teaspoons	46	you shipped pistols for you period	You ship teaspoons for you.
teaspoons	81	I should TSP for you period	I ship teaspoons for you.
teaspoons	105	I shipped teaspoons for you period	I ship teaspoons for you.
teens	23	you help things at school period	You help teens at school.
teens	26	I help things at school period	I help teens at school.
teens	33	we helped in the school period	We help teens at school.
teens	36	I helped in the school period	I help teens at school.
teens	73	we helped in the school period	We help teens at school.
teens	102	you helped in the school period	You help teens at school.
teeth	31	we help these problems period	We help teeth problems.
teeth	59	you help these problems period	You help teeth problems.
teeth	65	I help these problems period	I help teeth problems.
teeth	70	I help these problems period	I help teeth problems.
teeth	76	we have this problem period	We help teeth problems.
teeth	101	you help this problem period	You help teeth problems.

**APÊNDICE J – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 02 FEITA  
PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE**

<b>Palavra-alvo</b>	<b>Item</b>	<b>Transcrição youtube</b>	<b>Frase veículo</b>
keels	14	you got kill - 6 .	You got keels to fix.
keels	15	i got cute - 6 series	I got keels to fix.
keels	41	I got skills to fit	I got keels to fix.
keels	55	we got killed six dudes	We got keels to fix.
keels	58	you got guilty six spirit	You got keels to fix.
keels	108	my god gives to see period	We got keels to fix.
keep	16	I keep coffee here period	I keep coffee here.
keep	49	we keep calm here period	We keep coffee here.
keep	57	I keep coffee here cute	I keep coffee here.
keep	62	you keep coffee here give it	You keep coffee here.
keep	71	we keep coffee here period	You keep coffee here.
keep	90	we keep coffee year period	We keep coffee here.
keepers	2	I help keep it downstairs q and	I help keepers downstairs.
keepers	12	you help keepers notice to it	You help keepers downstairs.
keepers	22	we help keeper down sir serious	We help keepers downstairs.
keepers	56	you help people downstairs do it	You help keepers downstairs.
keepers	92	we help people not experience	We help keepers downstairs.
keepers	101	I help keep it downstairs period	I help keepers downstairs.
keyboards	19	we got keyboards to fix do it	We got keyboards to fix.
keyboards	79	you got keyboard to fix periods	You got keyboards to fix.
keyboards	85	I got keyboard to see do it	I got keyboards to fix.
keyboards	89	I got keyboard to fix periods	I got keyboards to fix.
keyboards	100	you got keyboard to fix period	You got keyboards to fix.
keyboards	107	we got keyboard to see period	We got keyboards to fix.
keys	21	I got killed doctor do	I got keys downstairs.
keys	30	you got keys doctors do	You got keys downstairs.
keys	40	I got kill doctors period	I got keys downstairs.
keys	50	we got killed doctors period	We got keys downstairs.
keys	52	we got kill doctor period	We got keys downstairs.
keys	87	you got kill dr. student	You got keys downstairs.
kiwi	8	build is QE for lunch - it	I eat kiwi for lunch.
kiwi	18	you will QE for lunch period	You eat kiwi for lunch.
kiwi	25	you is Q inference college group	I eat kiwi for lunch.
kiwi	26	we need QE for lunch do it	We eat kiwi for lunch.
kiwi	33	you eat chili for lunch period	You eat kiwi for lunch.
kiwi	39	we eat QE for lunch cute	We eat kiwi for lunch.
peace	51	we got this like a river to it	We got peace like a river.

peace	59	I got this like a reason serious	I got peace like a river.
peace	61	you got peace like a river view	You got peace like a river.
peace	63	you got peace like a river do it	You got peace like a river.
peace	81	we got this like a river being students	We got peace like a river.
peace	93	I got this like a riddle period	I got peace like a river.
peach	7	it is big Jim's yesterd	I ate peach yesterday.
peach	28	you a speech yesterday cute	You ate peach yesterday.
peach	80	we ate speech yesterday period	We ate peach yesterday.
peach	98	you wait pediatric experience	You ate peach yesterday.
peach	99	I ate speech yesterday to you	I ate peach yesterday.
peach	105	we ain't TGS period	We ate peach yesterday.
peaks	13	I walk pigs together do	I walk peaks together.
peaks	36	we walk pigs together cute	We walk peaks together.
peaks	73	you are big to get a student	You walk peaks together.
peaks	83	I won't dig together period	I walk peaks together.
peaks	86	I walk things together do it	You walk peaks together.
peaks	97	we walk things together period	We walk peaks together.
peanuts	24	I experienced yesterday do	I ate peanuts yesterday.
peanuts	38	we ate fellows yesterday cute	We ate peanuts yesterday.
peanuts	43	it is given pianos yesterday cute	I ate peanuts yesterday.
peanuts	53	you HDL of yesterday to it	You ate peanuts yesterday.
peanuts	65	you wait here not yesterday period	You ate peanuts yesterday.
peanuts	95	we hate this a a VHD energy every period	We ate peanuts yesterday.
people	47	you eat people doctors give it	You hit people downstairs.
people	60	you hate people downstairs spirit	You hit people downstairs.
people	64	we hit people downstairs period	We hit people downstairs.
people	69	we hit simple downstairs period	We hit people downstairs.
people	77	I hit people downstairs do it	I hit people downstairs.
people	96	I get people down there experience	I hit people downstairs.
pizza	1	you eat pizza for lunch period	You eat pizza for lunch.
pizza	29	I need pizza for lunch dude	I eat pizza for lunch.
pizza	68	we need pizza for lunch period	We eat pizza for lunch.
pizza	70	is pizza for lunch to it	I eat pizza for lunch.
pizza	91	we need is a for lunch period	We eat pizza for lunch.
pizza	104	you will be the for lunch period	You eat pizza for lunch.
teachers	3	you help teachers that school period	You help teachers at school.
teachers	44	I help teachers at school do	I help teachers at school.
teachers	66	I help teachers at school period	I help teachers at school.
teachers	74	you help teachers at school period	You help teachers at school.

teachers	102	we help teachers at school period	We help teachers at school.
teachers	106	we have teachers at school period	We help teachers at school.
team	5	we talk team fights period	We stop team fights.
team	10	you stopped in height cute	You stop team fights.
team	54	we stopped in fact stood	We stop team fights.
team	72	I stopped inside spirit	I stop team fights.
team	88	you stop King heights do it	You stop team fights.
team	103	I stopped in height period	I stop team fights.
teaser	11	we talked to the center to it	We talk teaser sentences.
teaser	20	I talked to the campus dude	I talk teaser sentences.
teaser	34	we talk get ourselves to use	We talk teaser sentences.
teaser	48	you talk to yourself to it	You talk teaser sentences.
teaser	78	I talk to the shelters period	I talk teaser sentences.
teaser	94	you talk to yourself as a student	You talk teaser sentences.
teaspoons	9	you ship TSP for you -	You ship teaspoons for you.
teaspoons	27	I ship default for you period	I ship teaspoons for you.
teaspoons	31	I shift useful for you period	I ship teaspoons for you.
teaspoons	35	will shift in fall for you do it	We ship teaspoons for you.
teaspoons	42	you ship did fall for you do	You ship teaspoons for you.
teaspoons	45	we shift is all for you do it	We ship teaspoons for you.
teens	6	we help change the school -	We help teens at school.
teens	17	I helped in the school give it	I help teens at school.
teens	23	you help things at school do	You help teens at school.
teens	32	we helped in the school do	We help teens at school.
teens	46	you have kings at school to it	You help teens at school.
teens	84	I helped King the school period	I help teens at school.
teeth	4	you help chief problems period	You help teeth problems.
teeth	37	we help this problem period	I help teeth problems.
teeth	67	you have these problems period	You help teeth problems.
teeth	75	I help this problem period	I help teeth problems.
teeth	76	we help to these problems do it	We help teeth problems.
teeth	82	we help this problem do it	We help teeth problems.

**APÊNDICE K – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 03 FEITA  
PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE**

Palavra-alvo	Item	Transcrição youtube	Frase veículo
keels	34	we got queues to fix period	We got keels to fix.
keels	46	you got killed to fix period	You got keels to fix.
keels	51	we got cute to fix period	We got keels to fix.
keels	52	I got killed to fix period	I got keels to fix.
keels	84	I got close to fix period	I got keels to fix.
keels	90	you got you got killed to fix period	You got keels to fix.
keep	14	I keep cut here period	I keep coffee here.
keep	32	we keep coffee here period	We keep coffee here.
keep	49	we keep copy here period	We keep coffee here.
keep	66	you keep copy here period	You keep coffee here.
keep	82	you keep coffee here period	You keep coffee here.
keep	99	I keep coughing year period	I keep coffee here.
keepers	6	you help keep your keepers downstairs period	You help keepers downstairs.
keepers	43	you helped keepers down tears period	You help keepers downstairs.
keepers	47	I help keepers downstairs period	I help keepers downstairs.
keepers	60	we helped keepers downstairs period	We help keepers downstairs.
keepers	100	we have keepers down there period	We help keepers downstairs.
keepers	102	I have keepers downstairs period	I help keepers downstairs.
keyboards	9	you got keyboard to fix period	You got keyboards to fix.
keyboards	24	I got keyboards to fix period	I got keyboards to fix.
keyboards	54	we got keyboard to fix period	We got keyboards to fix.
keyboards	55	(-)	We got keyboards to fix.
keyboards	61	you got keyboard to fit period	You got keyboards to fix.
keyboards	97	I got keyboards to fix period	I got keyboards to fix.
keys	15	you got keys downstairs period	You got keys downstairs.
keys	21	we got keys doctors period	We got keys downstairs.
keys	26	we got his dancers period	We got keys downstairs.
keys	33	you got kids downstairs period	You got keys downstairs.
keys	78	I got keys downstairs period	I got keys downstairs.
keys	108	I got heels down there's period	I got keys downstairs.
kiwi	17	you will kill me for lunch period	You eat kiwi for lunch.
kiwi	20	I kill me for lunch period	I eat kiwi for lunch.
kiwi	39	we 8qe for lunch period	We eat kiwi for lunch.
kiwi	45	I am Huy for lunch period	I eat kiwi for lunch.
kiwi	76	we need Kiwi for lunch period	We eat kiwi for lunch.
kiwi	94	you will queuing for lunch period	You eat kiwi for lunch.
peace	5	I've got these like a river perior	I got peace like a river.



peace	30	I got peace like a river period	I got peace like a river.
peace	41	you got this like River period	You got peace like a river.
peace	59	we got peace like a river period	We got peace like a river.
peace	74	we got peace like a river period	We got peace like a river.
peace	95	you got peace like a river period	You got peace like a river.
peach	2	I hate teach yesterday period	I ate peach yesterday.
peach	11	I beat yesterday period	I ate peach yesterday.
peach	42	you a speech yesterday period	You ate peach yesterday.
peach	53	we ate peach yesterday period	We ate peach yesterday.
peach	62	you will teach yesterday period	You ate peach yesterday.
peach	91	we in speech yesterday period	We ate peach yesterday.
peaks	58	you walk fit together period	You walk peaks together.
peaks	64	we want peaks together period	We walk peaks together.
peaks	77	I walk speaks together period	I walk peaks together.
peaks	86	I walk stick together period	I walk peaks together.
peaks	98	you will pick together period	You walk peaks together.
peaks	101	we walk pigs together period	We walk peaks together.
peanuts	10	I hate penis yesterday period	I ate peanuts yesterday.
peanuts	28	I a penis yesterday period	I ate peanuts yesterday.
peanuts	38	we ate peanuts yesterday period	We ate peanuts yesterday.
peanuts	63	you will finish yesterday period	You ate peanuts yesterday.
peanuts	83	we ate penis yesterday period	We ate peanuts yesterday.
peanuts	85	you wait finish yesterday period	You ate peanuts yesterday.
people	19	we hit people downstairs period	We hit people downstairs.
people	35	you hit people downstairs period	You hit people downstairs.
people	44	we hit people downstairs period	We hit people downstairs.
people	80	I give people downstairs period	I hit people downstairs.
people	88	I hit people downstairs period	I hit people downstairs.
people	104	you hit people downstairs period	You hit people downstairs.
pizza	3	we pizza for lunch period	We eat pizza for lunch.
pizza	8	i pizza for lunch period	I eat pizza for lunch.
pizza	18	you will pizza for lunch period	You eat pizza for lunch.
pizza	37	we pizza lunch period	We eat pizza for lunch.
pizza	65	you ate pizza for lunch period	You eat pizza for lunch.
pizza	75	I speak the for lunch period	I eat pizza for lunch.
teachers	16	you have teachers at school period	You help teachers at school.
teachers	56	we have teachers at call period	We help teachers at school.
teachers	89	you help teachers at school period	You help teachers at school.
teachers	92	we help teachers at home period	We help teachers at school.
teachers	93	I have teachers and period	I help teachers at school.

teachers	107	I have teachers at school period	I help teachers at school.
team	12	you stopped teamfight period	You stop team fights.
team	31	I stopped in fight period	I stop team fights.
team	68	a 15-5 period	I stop team fights.
team	71	we stopped in flight period	We stop team fights.
team	73	is stopped in flight period	You stop team fights.
team	96	with something fight period	We stop team fights.
teaser	7	you talk teaser sentences period	You talk teaser sentences.
teaser	25	I chop teaser sentences period	I talk teaser sentences.
teaser	27	I taught teaser sample period	I talk teaser sentences.
teaser	36	without teaser sentences period	We talk teaser sentences.
teaser	50	you talked inter sample period	You talk teaser sentences.
teaser	72	without teaser sample period	We talk teaser sentences.
teaspoons	40	you ship TSP for you period	You ship teaspoons for you.
teaspoons	57	we ship teaspoons for you period	We ship teaspoons for you.
teaspoons	69	I ship TSP for youth period	I ship teaspoons for you.
teaspoons	87	we shipped in spoons for you period	We ship teaspoons for you.
teaspoons	103	you shifting spoons for you period	You ship teaspoons for you.
teaspoons	105	as shifting from tin spoon for you period	I ship teaspoons for you.
teens	1	we helped things that whole period	We help teens at school.
teens	4	I helped in that call period	I help teens at school.
teens	70	you helped in the scope period	You help teens at school.
teens	79	we have things in that scope period	We help teens at school.
teens	81	I have things at school period	I help teens at school.
teens	106	you have seen that coal period	You help teens at school.
teeth	13	I helped teeth problems period	I help teeth problems.
teeth	22	you help teeth problem period	You help teeth problems.
teeth	23	(-)	You help teeth problems.
teeth	29	we have this problem period	We help teeth problems.
teeth	48	we have teeth problems period	We help teeth problems.
teeth	67	I help this problem period	I help teeth problems.

**APÊNDICE L – TRANSCRIÇÃO DAS AMOSTRAS DO SUJEITO 10 FEITA  
PELA FERRAMENTA DE TRANSCRIÇÃO DO YOUTUBE**

Palavra-alvo	Item	Transcrição youtube	Frase veículo
keels	33	we are killing the face feels	We got keels to fix.
keels	48	you got skills to fix period	You got keels to fix.
keels	62	you can choose to face spirit	You got keels to fix.
keels	73	I got skills to experience	I got keels to fix.
keels	88	you to feel spirit	We got keels to fix.
keels	94	I got skills to be experienced	I got keels to fix.
keep	17	I keep copy here	I keep coffee here.
keep	19	we keep copies here period	We keep coffee here.
keep	21	you keep copy here period	You keep coffee here.
keep	38	you keep copy here period	You keep coffee here.
keep	64	I did copy here build	I keep coffee here.
keep	76	we can copy year period	We keep coffee here.
keepers	15	I help people down there spirit	I help keepers downstairs.
keepers	18	we have people doctors spirits	We help keepers downstairs.
keepers	27	you have people doubt their purity	You help keepers downstairs.
keepers	29	we help keep yourself their spirit	We help keepers downstairs.
keepers	34	you have people down their spears	You help keepers downstairs.
keepers	44	I help people down their spirits	I help keepers downstairs.
keyboards	7	you don't keyboard to fix period	You got keyboards to fix.
keyboards	45	either keyboard to be experienced	You got keyboards to fix.
keyboards	67	because keyboard just experience	We got keyboards to fix.
keyboards	71	because people to this theory	We got keyboards to fix.
keyboards	90	I just want to take spirit	I got keyboards to fix.
keyboards	91	I got people to think spirit	I got keyboards to fix.
keys	12	swe have cheese down there spirit	We got keys downstairs.
keys	23	you can keep that their spirits	You got keys downstairs.
keys	39	you can kill down there till	You got keys downstairs.
keys	66	because he's downstairs spirit	We got keys downstairs.
keys	85	I just keep downstairs period	I got keys downstairs.
keys	93	I got killed out there spirit	I got keys downstairs.
kiwi	1	we here with her lunch period	We eat kiwi for lunch.
kiwi	10	I need to refer lunch periods	I eat kiwi for lunch.
kiwi	51	you will queue in for lunch series	You eat kiwi for lunch.
kiwi	54	with queue information	We eat kiwi for lunch.
kiwi	68	you is I did copy here build purely formal spirit	You eat kiwi for lunch.
kiwi	108	(-)	I eat kiwi for lunch.
peace	2	because peace like a river period	We got peace like a river.

peace	52	I got these logarithm theory because	I got peace like a river.
peace	53	this one little period	We got peace like a river.
peace	70	I just feel a little period	I got peace like a river.
peace	78	you can still under interference	You got peace like a river.
peace	104	you can feel like a recent period	You got peace like a river.
peach	37	you will each yesterday students	You ate peach yesterday.
peach	65	you a speech yesterday spirit	You ate peach yesterday.
peach	77	I teach gentle experience	I ate peach yesterday.
peach	89	this one little period	We ate peach yesterday.
peach	99	we will be G enter they feel	We ate peach yesterday.
peach	101	I am each the end of the period	I ate peach yesterday.
peaks	14	you will dig together period	You walk peaks together.
peaks	24	you won't fix together theories	You walk peaks together.
peaks	25	we all you need different experience stick together period	We walk peaks together.
peaks	30	we walk this together period	We walk peaks together.
peaks	47	I want this together period	I walk peaks together.
peaks	57	I want things together period	I walk peaks together.
peanuts	3	is thinner generally period	I ate peanuts yesterday.
peanuts	31	you will finish yesterday period	You ate peanuts yesterday.
peanuts	55	a business yesterday period	I ate peanuts yesterday.
peanuts	75	we are pinnate the internal period	We ate peanuts yesterday.
peanuts	87	we affinity activity because	We ate peanuts yesterday.
peanuts	106	you will business	You ate peanuts yesterday.
people	46	if people doubt their views	I hit people downstairs.
people	60	I had people doubt their spirit	I hit people downstairs.
people	69	you with people downstairs beauty	You hit people downstairs.
people	81	you make people doubt their experience	You hit people downstairs.
people	83	with people does their disputes	We hit people downstairs.
people	107	(-)	We hit people downstairs.
pizza	5	I'm pizza for lunch theory	I eat pizza for lunch.
pizza	22	(-)	You eat pizza for lunch.
pizza	26	you is season for lunch spirit	You eat pizza for lunch.
pizza	43	will be performance period	We eat pizza for lunch.
pizza	95	I see suffer love you	I eat pizza for lunch.
pizza	102	we beat the first series	We eat pizza for lunch.
teachers	16	we have teachers the school period	We help teachers at school.
teachers	35	i have teachers and school period	I help teachers at school.
teachers	50	I have teacher that full period	I help teachers at school.
teachers	72	we have featured at school period	We help teachers at school.

teachers	96	you have each other school period	You help teachers at school.
teachers	97	I have things at school period	You help teachers at school.
team	20	i starting my period	I stop team fights.
team	36	you stop in fact period	You stop team fights.
team	56	I thought inside pills	I stop team fights.
team	63	with nothing my fears	We stop team fights.
team	74	is boxing fight skills	You stop team fights.
team	79	we saw inside Spears	We stop team fights.
teaser	8	you talk to the sentence spirit	You talk teaser sentences.
teaser	9	without either sense of spirit	We talk teaser sentences.
teaser	13	without teasers and pirit	We talk teaser sentences.
teaser	40	you talk to the senses still	You talk teaser sentences.
teaser	58	I thought teaser sentences spirit	I talk teaser sentences.
teaser	84	a dog Jesus and the spirit	I talk teaser sentences.
teaspoons	11	you should people for you beauties	You ship teaspoons for you.
teaspoons	41	we seasons for you use	We ship teaspoons for you.
teaspoons	49	which is different for you do	We ship teaspoons for you.
teaspoons	59	I ship systems for you in period	I ship teaspoons for you.
teaspoons	86	you should be spoons for you Pierce	You ship teaspoons for you.
teaspoons	92	I should be able for you feel	I ship teaspoons for you.
teens	4	we helped in a school period	We help teens at school.
teens	6	you help things at school period	You help teens at school.
teens	32	you have teachers at school period	You help teens at school.
teens	82	I hope things at school period	I help teens at school.
teens	98	you have teachers at school period	I help teens at school.
teens	105	we have change the school period	We help teens at school.
teeth	28	we have these problems spirit	We help teeth problems.
teeth	42	you have hip problems views	You help teeth problems.
teeth	61	I have D strong spirit	I help teeth problems.
teeth	80	we have these problems beauty	We help teeth problems.
teeth	100	I have these problems period	I help teeth problems.
teeth	103	you have this problems period	You help teeth problems.